

# Multiplicative Approximations for Polynomial Optimization Over the Unit Sphere

Vijay Bhattiprolu\*    Mrinalkanti Ghosh<sup>†</sup>    Venkatesan Guruswami<sup>‡</sup>  
 Euiwoong Lee<sup>§</sup>    Madhur Tulsiani<sup>¶</sup>

November 21, 2016

## Abstract

We consider the following basic problem: given an  $n$ -variate degree- $d$  homogeneous polynomial  $f$  with real coefficients, compute a unit vector  $x \in \mathbb{R}^n$  that maximizes  $|f(x)|$ . Besides its fundamental nature, this problem arises in many diverse contexts ranging from tensor and operator norms to graph expansion to quantum information theory. The homogeneous degree 2 case is efficiently solvable as it corresponds to computing the spectral norm of an associated matrix, but the higher degree case is NP-hard.

In this work, we give multiplicative approximation algorithms for this problem. Our algorithms leverage the tractability of the degree 2 case, and output the best solution among a carefully constructed set of quadratic polynomials. They offer a trade-off between the approximation ratio and running time, which is governed by the number of quadratic problems we search over. Specifically, in  $n^{O(q)}$  time, we get an approximation within factor  $O_d((n/q)^{d/2-1})$  for arbitrary polynomials, and  $O_d((n/q)^{d/4-1/2})$  for polynomials with non-negative coefficients. The approximation guarantees are with respect to the optimum of the level- $q$  SoS SDP relaxation of the problem, which the algorithm rounds to a unit vector. We also consider the case when  $f$  is random with independent  $\pm 1$  coefficients, and prove that w.h.p the level- $q$  SoS solution gives a certificate within factor  $\tilde{O}_d((n/q)^{d/4-1/2})$  of the optimum.

We complement our algorithmic results with some polynomially large integrality gaps for  $d$ -levels of the SoS relaxation. For the random polynomial case, we show a gap of  $\Omega_d(n^{d/4-1/2})$ , which precisely matches the exponent of our upper bound, and shows the necessity of our  $\Omega(d)$  exponent in the approximation ratio for general polynomials. For polynomials with non-negative coefficients, we show an  $\tilde{\Omega}(n^{1/12})$  gap for the  $d = 4$  case.

To obtain our results, we develop general techniques which help analyze the approximation obtained by higher levels of the SoS hierarchy. We believe these techniques will also be useful in understanding polynomial optimization for other constrained settings.

---

\*Supported by NSF CCF-1422045 and CCF-1526092. vpb@cs.cmu.edu

<sup>†</sup>Supported by NSF CCF-1254044 mghosh@ttic.edu

<sup>‡</sup>Supported in part by NSF grant CCF-1526092. guruswami@cmu.edu

<sup>§</sup>Supported by a Samsung Fellowship, Simons award, and NSF CCF-1526092. euiwoonl@cs.cmu.edu

<sup>¶</sup>Supported by NSF CCF-1254044 madhurt@ttic.edu

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Our Results . . . . .	2
1.2	Related Previous and Recent Works . . . . .	4
<b>2</b>	<b>Preliminaries and Notation</b>	<b>6</b>
2.1	Polynomials . . . . .	6
2.2	Matrices . . . . .	7
<b>3</b>	<b>Overview of Proofs and Techniques</b>	<b>7</b>
3.1	Reduction to Optimization of Multi-linear Polynomials . . . . .	7
3.2	Relating $\Lambda(f)$ to $\ f\ _2$ for Multilinear Polynomials . . . . .	8
3.3	Improved Approximations via Folding . . . . .	9
3.4	Random Polynomials . . . . .	10
3.4.1	Upper Bounds . . . . .	10
3.4.2	Lower Bounds . . . . .	12
3.5	Lower Bounds Non-negative Coefficient Polynomials . . . . .	12
<b>4</b>	<b>Additional Preliminaries and the SoS Hierarchy</b>	<b>13</b>
4.1	Pseudoexpectations and Moment Matrices . . . . .	13
4.1.1	Constrained Pseudoexpectations . . . . .	14
4.2	Matrix Representations of Polynomials and $\Lambda(f)$ . . . . .	14
<b>5</b>	<b>Results for Polynomials in <math>\mathbb{R}_d[x]</math> and <math>\mathbb{R}_d^+[x]</math></b>	<b>15</b>
5.1	Detailed Overview . . . . .	15
5.1.1	Sufficient to analyse SoS-degree = Polynomial-degree regime . . . . .	15
5.1.2	SoS-degree = Polynomial-degree Regime and Multilinearity . . . . .	16
5.1.3	Decomposing Polynomials into Multilinear components . . . . .	16
5.1.4	Reduction to the Multilinear components . . . . .	17
5.2	Reduction to Multilinear Polynomials . . . . .	18
5.2.1	Upper Bounding $\Lambda(f)$ in terms of $\Lambda(F_{2\alpha})$ . . . . .	18
5.2.2	Lower Bounding $\ f\ _2$ in terms of $\ F_{2\alpha}\ _2$ (non-negative coefficients) . . . . .	18
5.2.3	Lower Bounding $\ f\ _2$ in terms of $\ F_{2\alpha}\ _2$ (general case) . . . . .	19
5.3	$(n/q)^{d/4}$ -Approximation for Non-negative Coefficient Polynomials . . . . .	22
5.4	$(n/q)^{d/2}$ Approximation for General Polynomials . . . . .	23
<b>6</b>	<b>Approximating 2-norms via Folding</b>	<b>24</b>
6.1	Preliminaries . . . . .	24
6.2	Reduction to Multilinear Folded Polynomials . . . . .	24
6.2.1	Relating $\Lambda_C(f)$ to $\Lambda_C(F_{2\alpha})$ . . . . .	25
6.3	Relating Evaluations of $f$ to Evaluations of $F_{2\alpha}$ . . . . .	25

6.4	Bounding $\Lambda_C()$ of Multilinear Folded Polynomials . . . . .	26
6.4.1	General Folds: Bounding $\Lambda()$ in terms of $\Lambda()$ of the "worst" fold . . . . .	26
6.4.2	Non-Negative Coefficient Folds: Relating SoS Value to the SoS Value of the $d_1/2$ -collapse . . . . .	27
6.5	$(n/q)^{d/4-1/2}$ -Approximation for Non-negative Coefficient Polynomials . . . . .	28
6.6	$(n/q)^{d/2-1}$ -Approximation for General Polynomials . . . . .	30
6.7	Algorithms . . . . .	31
6.7.1	Non-negative coefficient polynomials . . . . .	31
6.7.2	General Polynomials . . . . .	31
<b>7</b>	<b>Random Polynomials</b>	<b>32</b>
7.1	Upper Bounds . . . . .	32
7.2	Lower Bounds . . . . .	35
7.2.1	Wigner Moment Matrix . . . . .	35
7.2.2	Final Construction . . . . .	38
<b>8</b>	<b>Lower Bounds for Non-negative Coefficient Polynomials</b>	<b>41</b>
8.1	Upper Bound on $\ f\ _2$ . . . . .	41
8.1.1	Discretization . . . . .	41
8.1.2	Bounding Edge Density . . . . .	42
8.1.3	Bounding Triangle Density . . . . .	43
8.1.4	Bounding 4-clique Density . . . . .	45
8.2	Lower Bound on $\Lambda(f)$ . . . . .	46
<b>A</b>	<b>Oracle Lower Bound</b>	<b>54</b>

# 1. Introduction

We study the problem of optimizing homogeneous polynomials over the unit sphere. Formally, given an  $n$ -variate degree- $d$  homogeneous polynomial  $f$ , the goal is to compute

$$\|f\|_2 := \sup_{\|x\|=1} |f(x)| \quad (1.1)$$

When  $f$  is a homogeneous polynomial of degree 2, this problem is equivalent computing the spectral norm of an associated symmetric matrix  $M_f$ . For higher degree  $d$ , it defines a natural higher-order analogue of the eigenvalue problem for matrices. The problem also provides an important test case for the development of new spectral and semidefinite programming (SDP) techniques, and techniques developed in the context of this problem have had applications to various other constrained settings [HLZ10, Lau09, Las09].

Besides being a natural and fundamental problem in its own right, it has connections to widely studied questions in many other areas. In quantum information theory [BH13, BKS14], the problem of computing the optimal success probability of a protocol for Quantum Merlin-Arthur games can be thought of as optimizing certain classes of polynomials over the unit sphere. The problem of estimating the  $2 \rightarrow 4$  norm of an operator, which is equivalent to optimizing certain homogeneous degree-4 polynomials over the sphere, is known to be closely related to the Small Set Expansion Hypothesis (SSEH) and the Unique Games Conjecture (UGC) [BBH<sup>+</sup>12, BKS14]. The polynomial optimization problem is also very relevant for natural extensions of spectral problems, such as low-rank decomposition and PCA, to the case of tensors [BKS15, GM15, MR14, HSS15]. Frieze and Kannan [FK08] (see also [BV09]) also established a connection between the problem of approximating the spectral norm of a tensor (or equivalently, computing  $\|f\|_2$  for a polynomial  $f$ ), and finding planted cliques in random graphs.

The problem of polynomial optimization has been studied<sup>1</sup> over various compact sets [Las09, DK08], and is natural to ask how well polynomial time algorithms can *approximate* the optimum value over a given compact set (see [DK08] for a survey). While the maximum of a degree- $d$  polynomial over the simplex admits a PTAS for every fixed  $d$  [DKLP06], the problem of optimizing even a degree 3 polynomial over the hypercube does not admit any approximation better than  $2^{(\log n)^{1-\epsilon}}$  (for arbitrary  $\epsilon > 0$ ) assuming NP cannot be solved in time  $2^{(\log n)^{O(1)}}$  [HV04].

The approximability of polynomial optimization on the sphere is poorly understood in comparison. It is known that the maximum of a degree- $d$  polynomial can be approximated within a factor  $n^{d/2-1}$  in polynomial time [HLZ10, So11]. On the hardness side, Nesterov [Nes03] gave a reduction from Maximum Independent Set to optimizing a homogeneous cubic polynomial over  $S^{n-1}$ . Formally, given a graph  $G$ , there exists a homogeneous cubic polynomial  $f(G)$  such that  $\sqrt{1 - \frac{1}{\alpha(G)}} = \max_{\|x\|=1} f(x)$ . Combined with the hardness of Maximum Independent Set [Hås96], this rules out an FPTAS for optimization over the unit sphere. Assuming the Exponential Time Hypothesis, Barak et al. [BBH<sup>+</sup>12] proved that computing  $2 \rightarrow 4$  norm of a matrix, a special case when  $f$  is a degree-4 homogeneous polynomial, is hard to approximate within a factor  $\exp(\log^{1/2-\epsilon}(n))$  for any  $\epsilon > 0$ .

One of the popular approaches to attack the above problem, called the *Sum of Squares Hierarchy* (SoS), proceeds by replacing a system of non-negativity constraints by a suitable sum of squares

---

<sup>1</sup>In many cases the problem studied, is not to maximize  $|f|$ , but just  $f(x)$ . While the two problems are equivalent for homogeneous polynomials of odd degree, some subtle issues arise when considering polynomials of even degree. We compare the two notions later.

decomposition. Algorithms based on this framework are parametrized by the degree  $q$  of their SoS decomposition. Optimization over  $\mathbb{S}^{n-1}$  via SoS has been given attention in the optimization community, where for a fixed number of variables  $n$  and degree  $d$  of the polynomial, it is known that the estimates produced by the SoS hierarchy get arbitrarily close to the true optimal solution as  $q$  increases. We refer the reader to the recent work of Doherty and Wehner [DW12] and de Klerk, Laurent, and Sun [dKLS14] and references therein for more information on convergence results. By using semidefinite programming, these algorithms run in time  $n^{O(q)}$ , which is polynomial for constant  $q$ . Unfortunately, known convergence results often give a non-trivial bound only when the degree parameter  $q$  is linear in  $n$ .

In computer science, much attention has been given to the regime  $d \leq q \ll n$ , so that the resulting algorithm runs in at most subexponential time. In addition to the  $n^{d/2-1}$  approximation for general polynomials [HLZ10, So11], approximation guarantees have been proved for several special cases including  $2 \rightarrow q$  norms [BBH<sup>+</sup>12], non-negative polynomials [BKS14], and some polynomials that arise in quantum information theory [BH13]. As such, there is considerable interest in tightly characterizing the approximation guarantee of the SoS hierarchy on many of these special cases.

Several of the improved guarantees above, obtained using the SoS hierarchy, suffer an *additive* loss depending on the *spectral norm* for the polynomial  $f$ , defined (by [BKS14]) as

$$\|f\|_{sp} := \inf \left\{ \|M\|_2 \mid M \in \mathbb{S}^{n^{d/2}}(\mathbb{R}), (x^{\otimes d/2})^T \cdot M \cdot x^{\otimes d/2} = f(x) \ \forall x \in \mathbb{R}^n \right\},$$

for a degree- $d$  homogeneous polynomial  $f$ , when  $d$  is even. While  $\|f\|_{sp}$  is bounded for some cases of interest, such as the polynomials corresponding to problems in quantum information theory, it can be quite large in general. In fact, (a variant of) the spectral norm can be shown to be equivalent the value of the SDP relaxation obtained by the  $d$  levels of the SoS hierarchy, and thus an algorithm obtaining an additive loss of  $\varepsilon \cdot \|f\|_{sp}$  using  $q(\varepsilon)$  can also be viewed as obtaining a multiplicative approximation guarantee (see Section 4). The dependence of  $q$  on  $\varepsilon$  can be viewed as describing a rate of improvement of the approximation factor with  $q$ .

In this paper, we study *multiplicative* approximation guarantees for polynomial optimization. We show that this can be directly analyzed for arbitrary degree- $d$  polynomials, as well as several special cases, including some where additive approximations are not yet known. We develop general techniques which help analyze the approximation obtained by higher levels of the SoS hierarchy for arbitrary polynomials, as well as give improved results for special cases such as polynomials with non-negative coefficients and random polynomials, when no apriori bound is known on  $\|f\|_{sp}$ . We believe these techniques will also be useful in understanding polynomial optimization for other constrained settings.

## 1.1. Our Results

We study a variant of the quantity  $\|f\|_{sp}$  defined by Barak et al. [BKS14]. For a homogeneous polynomial  $f$  of even degree  $d$ , we define the quantity

$$\Lambda(f) := \inf \left\{ \sup_{\|z\|=1} z^T M z \mid M \in \mathbb{S}^{n^{d/2}}(\mathbb{R}), (x^{\otimes d/2})^T \cdot M \cdot x^{\otimes d/2} = f(x) \ \forall x \in \mathbb{R}^n \right\}.$$

Note that while  $\|f\|_{sp}$  considers the spectral norm of  $M$ ,  $\Lambda(f)$  considers only the maximum eigenvalue. It is well known (see [Lau09])  $\Lambda(f)$  is equal to the dual value of the degree- $d$  SoS relaxation

for the problem of maximizing  $f$ . It is also known that strong duality holds for the case of optimization on the sphere and thus the two values are equal, and  $\Lambda(f)$  can be computed in time  $n^{O(d)}$ .

It is clear from the above definition that  $\Lambda(f)$  is a relaxation for  $\sup_{\|x\|=1} f(x)$ . Moreover, if  $f$  satisfies  $\|f\|_2 = \sup_{\|x\|=1} f(x)$ , then  $\Lambda(f)$  is also a relaxation for  $\|f\|_2$ . When this is not the case, we still have that  $(\Lambda(f^2))^{1/2}$  is a relaxation for  $\|f\|_2$ . Moreover, for any  $q$  divisible by  $2d$ , we have

$$\|f\|_2 \leq \left( \Lambda(f^{q/d}) \right)^{d/q} \leq (\Lambda(f^2))^{1/2} \leq \|f\|_{sp},$$

since if  $(x^{\otimes d})^T \cdot M \cdot x^{\otimes d} = f^2(x)$  for all  $x$ , then  $(x^{\otimes q/2})^T \cdot M^{\otimes q/d} \cdot x^{\otimes q} = f^q(x)$  for all  $x$ . We prove the following result for a general polynomial of degree  $d$ .

**Theorem 1.1.** *Let  $f$  be an  $n$  variate homogeneous polynomial of degree- $d$ , and let  $q$  be an integer divisible by  $2d$ . Then,*

$$\left( \Lambda(f^{q/d}) \right)^{d/q} \leq O_d\left((n/q)^{d/2-1}\right) \cdot \|f\|_2.$$

In fact, our proof is algorithmic and we prove that there is a deterministic algorithm, which for a given  $f$ , outputs an  $x$  such that

$$|f(x)| \geq \Omega_d\left((n/q)^{-(d/2-1)}\right) \cdot \left( \Lambda(f^{q/d}) \right)^{d/q}.$$

In the statements above and the rest of this section,  $O_d(\cdot)$  and  $\Omega_d(\cdot)$  notations hide  $2^{O(d)}$  factors. We prove similar theorems for various classes of degree- $d$  polynomials, with better bounds on the approximation ratio, as given in the table below.

Class of degree- $d$ polynomials	Approximation by $q$ levels	Lower bound on $\Lambda(f) / \ f\ _2$
Arbitrary	$O_d\left(\left(\frac{n}{q}\right)^{d/2-1}\right)$	$\Omega_d\left(\left(\frac{n}{d}\right)^{d/4-1/2}\right)$
Random	$\tilde{O}_d\left(\left(\frac{n}{q}\right)^{d/4-1/2}\right)$	$\Omega_d\left(\left(\frac{n}{d}\right)^{d/4-1/2}\right)$
Polynomials with non-neg. coefficients	$O_d\left(\left(\frac{n}{q}\right)^{d/4-1/2}\right)$	$\tilde{\Omega}(n^{1/12})$ for $d = 4$

**Remark 1.2.** *In cases where  $\|f\|_2 = \sup_{\|x\|=1} f(x)$  (such as when  $d$  is odd or  $f$  has non-negative coefficients), the above result holds whenever  $q$  is even and divisible by  $d$ , instead of  $2d$ .*

**Maximizing  $|f(x)|$  vs.  $f(x)$ .** Let  $\lambda_{\max}(f)$  denote  $\sup_{\|x\|=1} f(x)$ . Note that for polynomials with odd-degree, we have  $\|f\|_2 = \lambda_{\max}(f)$ . When the degree is even, a multiplicative approximation for  $\lambda_{\max}(f)$  is not possible since  $\lambda_{\max}(f)$  may be 0 or even negative. Moreover, even when  $\lambda_{\max}(f)$  is positive, any constructive multiplicative approximation of  $\lambda_{\max}(f)$  with a factor (say)  $B$ , can be turned into a  $1 + \varepsilon$  approximation by considering  $f' = f - C \cdot \|x\|_2^d$ , for  $C = (1 - \varepsilon) \cdot \lambda_{\max}(f)$  (one can use binary search on the values of  $C$  and use the solution give by the constructive algorithm to check).

An alternate notion considered in the literature [HLZ10, So11] is that of relative approximation where one bounds the ratio  $(\Lambda - \lambda_{\min}(f)) / (\lambda_{\max}(f) - \lambda_{\min}(f))$  (known as a relative approximation), where  $\Lambda$  is the estimate by the algorithm, and  $\lambda_{\min}(f)$  is defined analogously to  $\lambda_{\max}(f)$ . While this notion is stronger than approximating  $\|f\|_2$  in some cases, one can use a shift of  $f$  as in the example above (by  $C \cdot \lambda_{\min}(f)$ ) to obtain a relative approximation unless  $|\lambda_{\max}(f) - \lambda_{\min}(f)| / |\lambda_{\min}(f)| = n^{-\omega(1)}$ .



## Lower Bounds

We also study lower bounds on the quality approximation for  $\lambda_{\max}(f)$  provided by  $\Lambda(f)$ .

We first consider the case of random polynomials, which also provides a lower bound for the case of arbitrary polynomials. Here, we consider the model of random polynomials produced by random tensors or order  $d$ . Let  $T \in \{\pm 1\}^{[n]^d}$  be a random order- $d$  tensor where for each  $(i_1, \dots, i_d) \in [n]^d$ ,  $T(i_1, \dots, i_d)$  is an independent Rademacher ( $\pm 1$ ) random variable. We consider the polynomial  $f_T(x) = \langle T, x^{\otimes d} \rangle$ .

Given a random polynomial  $f_T$  of degree  $d$  as above, it is known that  $\|f\|_2 = O(\sqrt{n \log d})$  with high probability over the choice of coefficients. (see [TS14]).<sup>2</sup> We prove the following lower bound on  $\Lambda(f)$ .

**Theorem 1.3.** *Let  $T \in \{\pm 1\}^{[n]^d}$  be a random tensor of order  $d$  and let  $f_T$  be the corresponding homogeneous polynomial of degree  $d$ . Then, with high probability over the choice of  $T$ , we have*

$$\Lambda(f_T) \geq \Omega_d\left(\left(\frac{n}{d}\right)^{d/4}\right) = \Omega_d\left(\left(\frac{n}{d}\right)^{d/4-1/2}\right) \cdot \|f_T\|_2.$$

We also give a new lower bound construction for the case of non-negative polynomials, using subgraph counts in a random graph. To the best of our knowledge, the only previous lower bound for this problem, was known through Nesterov's reduction [DK08], which can be used to show that  $\Lambda(f) \geq (1 + \varepsilon)\lambda_{\max}(f)$  for some fixed constant  $\varepsilon > 0$ . We give the following polynomial lower bound.

**Theorem 1.4.** *There exists an  $n$  variate degree-4 homogeneous polynomial  $f$  with non-negative coefficients such that*

$$\|f\|_2 = \tilde{O}(n^{1/12}) \quad \text{and} \quad \Lambda(f) = \Omega(n^{1/6}).$$

## 1.2. Related Previous and Recent Works

Polynomial optimization is a vast area with several previous results. Below, we collect the results most relevant for comparison with the ones in this paper. Please see the excellent monographs [Lau09, Las09] for a survey.

### Arbitrary Polynomials

For general homogeneous polynomials of degree- $d$ , an  $O_d(n^{d/2-1})$  approximation was given by He et al. [HLZ10], which was later improved to  $O_d((n/\log n)^{d/2-1})$  by So [So11]. The convergence of SDP hierarchies for polynomial optimization was analyzed by Doherty and Wehner [DW12]. However, their result only applies to relaxations given by  $\Omega(n)$  levels of the SoS hierarchy. Thus, our results can be seen as giving an interpolation between the polynomial time algorithms obtained by [HLZ10, So11] and the exponential time algorithms given by  $\Omega(n)$  levels of SoS, although the bounds obtained by [DW12] are tighter (by a factor of  $2^{O(d)}$ ) for  $q = \Omega(n)$  levels.

We believe a tradeoff between running time and approximation quality similar to ours can also be obtained by considering the tradeoffs for the results of Brieden et al. [BGK<sup>+</sup>01] used by So

---

<sup>2</sup> [TS14] proves this for injective norm and it is not hard to specialize their proof for the two-norm case and obtain a better bound of  $\sqrt{n \log d}$

[So11]. However, to the best of our knowledge, this is not published. In particular, So uses the techniques of Khot and Naor [KN08] to reduce degree- $d$  polynomial optimization to  $d - 2$  instances of the problem of optimizing the  $\ell_2$  diameter of a convex body. This is solved by [BGK<sup>+</sup>01], who give a  $O((n/k)^{1/2})$  approximation in time  $2^k \cdot n^{O(1)}$ . We believe this can be combined with proof of So, to yield a  $O_d((n/q)^{d/2-1})$  approximation in time  $2^q$ . We note here that the method of Khot and Naor [KN08] cannot be improved further (up to polylog) for the case  $d = 3$  – see Appendix A. Our results for the case of arbitrary polynomials show that similar bounds can also be obtained by a very generic algorithm given by the SoS hierarchy. Moreover, the general techniques developed here are also relevant for other cases discussed below where no alternate proofs are available.

### Polynomials with Non-negative Coefficients

The case of polynomials with non-negative coefficients was considered by Barak et al. [BKS14] who proved that the relaxation obtained by  $\Omega(d^3 \cdot \log n / \varepsilon^2)$  levels of the SoS hierarchy provides an  $\varepsilon \cdot \|f\|_{sp}$  additive approximation to the quantity  $\|f\|_2$ .

Their results can be phrased as showing that a relaxation obtained by  $q$  levels of the SoS hierarchy give an approximation ratio of

$$\left( \frac{d^3 \cdot \log n}{q} \right)^{1/2} \cdot \frac{\|f\|_{sp}}{\|f\|_2}.$$

Our results show that the relaxation obtained by  $q$  levels of SOS gives an  $(n/q)^{d/4-1/2}$  multiplicative approximation to  $\|f\|_2$  which has a better dependence on  $q$ . However, the two results are incomparable since the worst case gap between  $\|f\|_{sp}$  and  $\|f\|_2$  is not completely understood. In the case of arbitrary polynomials, this gap is at least as large as  $\Omega_d((n/d)^{d/4-1/2})$  (which is implied by the degree- $d$  SOS lower bounds on degree- $d$  random polynomials obtained in this paper). As for the case of non-negative coefficient polynomials, to our knowledge the only polynomial sized gap is  $n^{1/12}$  which is established in the present work.

### Random Polynomials

Montanari and Richard [MR14] presented a  $n^{O(d)}$ -time algorithm that can certify that the optimal value is at most  $O(n^{\frac{[d/2]}{2}})$  with high probability. Hopkins, Shi, and Steurer [HSS15] improved it to  $O(n^{\frac{d}{4}})$  with the same running time. They also asked how many levels of SoS are required to certify a bound of  $n^{3/4-\delta}$  for  $d = 3$ .

Our analysis asymptotically improves the aforementioned bound when  $q$  is growing with  $n$ , and we prove an essentially matching lower bound (but only for the case  $q = d$ ). Secondly, we consider the case when  $d$  is fixed, and give improved results for the performance of degree- $q$  SoS (for large  $q$ ), thus answering in part, a question posed by Hopkins, Shi and Steurer [HSS15].

Raghavendra, Rao, and Schramm [RRS16] have independently and concurrently obtained similar (slightly weaker) results as ours, for random degree- $d$  polynomials. They also generalized their techniques to give SoS upper bounds for certifying the norms of sparse random tensors which allowed them to give tight SoS upper bounds for refuting random instances of MAX k-LIN. Hopkins, Kothari, and Potechin [HKP16b] recently proved the following theorem, almost matching our upper bound on random polynomials.



**Theorem 1.5.** *Let  $f$  be a degree- $d$  polynomial with i.i.d. Gaussian coefficients. Then for every constant  $\epsilon > 0$ , with high probability over  $f$ , there is a degree  $n^\epsilon$  pseudodistribution with  $\tilde{\mathbb{E}}[f(x)] \geq n^{d/4 - O(d\epsilon)}$ , which satisfies  $\|x\|^2 = 1$ .*

## 2. Preliminaries and Notation

### 2.1. Polynomials

We use  $\mathbb{R}_d[x]$  to denote the set of all homogeneous polynomials of degree (exactly)  $d$ . Similarly,  $\mathbb{R}_d^+[x]$  is used to denote the set of polynomials with non-negative coefficients. All polynomials considered in this paper will be  $n$ -variate (with  $x$  denoting the set of  $n$  variables  $x_1, \dots, x_n$ ) unless otherwise stated.

A multi-index is defined as sequence  $\alpha \in \mathbb{N}^n$ . We use  $|\alpha|$  to denote  $\sum_{i=1}^n \alpha_i$  and  $\mathbb{N}_d^n$  (resp.  $\mathbb{N}_{\leq d}^n$ ) to denote the set of all multi-indices  $\alpha$  with  $|\alpha| = d$  (resp.  $|\alpha| \leq d$ ). Thus, a polynomial  $f \in \mathbb{R}_d[x]$  can be expressed in terms of its coefficients as

$$f(x) = \sum_{\alpha \in \mathbb{N}_d^n} f_\alpha \cdot x^\alpha,$$

where  $x^\alpha$  is used to denote the monomial corresponding to  $\alpha$ . A polynomial is multilinear if  $\alpha \leq \mathbb{1}$  whenever  $f_\alpha \neq 0$ , where  $\mathbb{1}$  denotes the multi-index  $1^n$ . We use the notation  $\alpha^r$  to denote the vector  $(\alpha_1^r, \dots, \alpha_n^r)$  for  $r \in \mathbb{R}$ . In general, with the exception of absolute-value, any scalar function when applied to a vector/multi-index returns the vector obtained by applying the function entry-wise. We also use  $\circ$  to denote the Hadamard (entry-wise) product of two vectors.

To get the constant terms in the exponent of our results, we will need to extract the “quadratic part” of a given polynomial, and use the fact that eigenvalue problems are easy for quadratic polynomials. We thus define, the following polynomials where the coefficients themselves may be polynomials (in the same variables).

**Definition 2.1 (Folded Polynomials).** *A degree- $(d_1, d_2)$  folded polynomial  $f \in (\mathbb{R}_{d_2}[x])_{d_1}[x]$  is defined to be a polynomial of the form*

$$f(x) = \sum_{\alpha \in \mathbb{N}_{d_1}^n} \bar{f}_\alpha(x) \cdot x^\alpha,$$

where each  $\bar{f}_\alpha(x) \in \mathbb{R}_{d_2}[x]$  is a homogeneous polynomial of degree  $d_2$ . Folded polynomials over  $\mathbb{R}^+$  are defined analogously.

- We refer to the polynomials  $\bar{f}_\alpha$  as the folds of  $f$  and the terms  $x^\alpha$  as the monomials in  $f$ .
- A folded polynomial can also be used to define a degree  $d_1 + d_2$  polynomial by multiplying the monomials with the folds (as polynomials in  $\mathbb{R}[x]$ ). We refer to this polynomial in  $\mathbb{R}_{d_1+d_2}[x]$  as the unfolding of  $f$ , and denote it by  $U(f)$ .
- For a degree  $(d_1, d_2)$ -folded polynomial  $f$  and  $r \in \mathbb{N}$ , we take  $f^r$  to be a degree- $(r \cdot d_1, r \cdot d_2)$  folded polynomial, obtained by multiplying the folds as coefficients.

We define the following operation (and it’s folded counterpart) which in the case of a multilinear polynomial corresponds (upto scaling) to the sum of a row of the SOS symmetric matrix representation of the polynomial.

**Definition 2.2 (Collapse).** Let  $f \in \mathbb{R}_d[x]$  be a polynomial. The  $k$ -collapse of  $f$ , denoted as  $C_k(f)$  is the degree  $d - k$  polynomial defined with coefficients

$$g_\gamma = \sum_{\alpha \in \mathbb{N}_k^n} f_{\gamma+\alpha}.$$

For a degree- $(d_1, d_2)$  folded polynomial  $f$ , we define  $C_k(f)$  similarly as the degree- $(d_1 - k, d_2)$  folded polynomial  $g$  given by

$$g = \sum_{\gamma \in \mathbb{N}_{d_1-k}^n} \bar{g}_\gamma(x) \cdot x^\gamma \quad \text{where} \quad \bar{g}_\gamma = \sum_{\alpha \in \mathbb{N}_k^n} \bar{f}_{\gamma+\alpha}.$$

## 2.2. Matrices

For  $k \in \mathbb{N}$ , we will consider matrices  $M \in \mathbb{R}^{n^k \times n^k}$ . All matrices considered in this paper should be taken to be symmetric (unless otherwise stated). We index entries of the matrix  $M$  as  $M[I, J]$  by tuples  $I, J \in [n]^k$ .

A tuple  $I = (i_1, \dots, i_k)$  naturally corresponds to a multi-index  $\alpha(I) \in \mathbb{N}_k^n$  with  $|\alpha(I)| = k$ . For a tuple  $I \in [n]^k$ , we define  $\mathcal{O}(I)$  the set of all tuples  $J$  which correspond to the same multi-index i.e.,  $\alpha(I) = \alpha(J)$ . Thus, each  $\alpha$  corresponds to an equivalence class in  $[n]^k$ . We also use  $\mathcal{O}(\alpha)$  to denote the class of all tuples corresponding to  $\alpha$ .

Note that a matrix of the form  $(x^{\otimes k})(x^{\otimes k})^T$  has many additional symmetries, which are also present in solutions to programs given by the SoS hierarchy. To capture this, a matrix  $M$  which satisfies  $M[I, J] = M[K, L]$  whenever  $\alpha(I) + \alpha(J) = \alpha(K) + \alpha(L)$  is referred to as SoS-symmetric.

## 3. Overview of Proofs and Techniques

### 3.1. Reduction to Optimization of Multi-linear Polynomials

One of the main techniques we develop in this work, is a way of reducing the optimization problem for general polynomials to that of multi-linear polynomials, which *does not increase the number of variables*. While general techniques for reduction to the multi-linear case have been widely used in the literature [KN08, HLZ10, So11], these reduce the problem to optimizing a multi-linear polynomial in  $n \cdot d$  variables for a polynomial  $f$  of degree  $d$ . Since we are interested in the improvement in approximation obtained by considering  $f^{q/d}$  for a large  $q$ , applying these would yield a multi-linear polynomial in  $n \cdot q$  variables. For our analysis, this increase in variables exactly cancels the advantage we obtain by considering  $f^{q/d}$  instead of  $f$  (i.e., the advantage obtained by using  $q$  levels of the SoS hierarchy).

For a given polynomial  $f$  of degree- $d$ , we assume below that  $\|f\|_2 = \sup_{\|x\|=1} f(x)$  and  $d$  is even. If this not the case, we reduce to this case by considering  $g = f^2$ . We can uniquely represent a homogeneous polynomial  $f$  of degree  $d$  as

$$f(x) = \sum_{|\alpha| \leq d/2} F_{2\alpha}(x) \cdot x^{2\alpha},$$

where each  $F_{2\alpha}$  is a multi-linear polynomial. We reduce the problem to optimizing  $\|F_{2\alpha}\|_2$  for each of the polynomials  $F_{2\alpha}$ . We discuss later that  $\Lambda(F_\alpha)$  is easy to relate to  $\|F_{2\alpha}\|_2$  for multi-linear polynomials  $F_{2\alpha}$ , since the matrices arising in the definition of  $\Lambda(F_\alpha)$  can be highly symmetrized

in this case. At a high level, the aforementioned decomposition is motivated by the idea that when optimizing the multilinear components of a polynomial, one can either exploit that there is a high degree of multilinearity (when  $|\alpha|$  is small) and obtain better dependence on the exponent of  $d$ , or one can exploit that the problem is effectively of lower degree (when  $|\alpha|$  is large) and obtain a better dependence on the exponent of  $n$  in the approximation.

Note that if each  $F_{2\alpha}$  is constant, i.e. if  $f$  only has terms of even degree in each variable, then by substituting  $y_i = x_i^2$  for each  $i$ , we can reduce to optimizing a degree- $(d/2)$  over the simplex, which admits a PTAS. For arbitrary multi-linear polynomials  $F_{2\alpha}$ , we show that

$$\Lambda(f) \leq \left(\frac{d}{2} + 1\right) \cdot \max_{|\alpha| \leq d/2} \frac{\Lambda(F_{2\alpha})}{|\mathcal{O}(\alpha)|} \quad \text{and} \quad \|f\|_2 \geq \frac{\|F_{2\alpha}\|_2}{2^{O(d)} \cdot |\mathcal{O}(\alpha)|} \quad \forall \alpha.$$

The first inequality follows by noticing that any matrix  $M_f$  such that  $(x^{\otimes(d/2)})^T \cdot M_f \cdot x^{\otimes(d/2)}$  for all  $x$  (called a matrix representation of  $f$ ) can be written as a sum of matrices  $M_{t,f}$  for each  $t \leq d/2$ , each of which is block-diagonal matrix with blocks corresponding to matrix representations of the polynomials  $M_{F_{2\alpha}}$  for each  $\alpha$  with  $|\alpha| = 2t$ .

For the second inequality, first consider the case when the coefficients of  $f$ , and hence also of  $F_{2\alpha}$  for each  $\alpha$ , are non-negative. If  $x^*$  is the optimizer of  $F_{2\alpha}$ , then it is easy to see that  $x^* \geq 0$  and

$$y = x^* + \sum_{i=1}^n \frac{\sqrt{\alpha_i}}{\sqrt{|\alpha|}} \cdot e_i = x^* + \frac{\sqrt{\alpha}}{\sqrt{|\alpha|}}$$

achieves a value greater than  $\|F_{2\alpha}\|_2 / (2^{O(d)} \cdot |\mathcal{O}(\alpha)|)$ . Here,  $\{e_i\}_{i \in [n]}$  denote the standard basis for  $\mathbb{R}^n$  and  $\sqrt{\alpha}$  denotes the vector  $(\sqrt{\alpha_1}, \dots, \sqrt{\alpha_n})$ .

For the case of general polynomials, this part requires significantly more care. We consider  $\|f\|_2^c = \sup_{z \in \mathbb{C}^n, \|z\|=1} |f(z)|$  and show that it is within a factor  $2^{O(d)}$  of  $\|f\|_2$ . We then give a lower bound for  $\|f\|_2^c$  in terms of  $\|F_{2\alpha}\|_2$  by considering the random vector

$$z = \Xi \cdot \sum_{i=1}^n \frac{b_i \cdot x_i^*}{2\alpha_i + 1} + \sum_{i=1}^n \omega_i \cdot \frac{\sqrt{\alpha_i}}{\sqrt{|\alpha|}} \cdot e_i = \Xi \cdot b \circ \frac{1}{2\alpha + \mathbb{1}} \circ x^* + \omega \circ \frac{\sqrt{\alpha}}{\sqrt{|\alpha|}},$$

where  $\Xi$  is a random  $(d - 2|\alpha| + 1)^{th}$  root of unity, each  $\omega_i$  is a random  $(2\alpha_i + 1)^{th}$  root of unity, and  $b$  is a vector of  $n$  independent  $\{0, 1\}$ -valued variables with  $\mathbb{E}[b_i] = p$  for each  $i$ . We use  $\circ$  to denote the Hadamard (entry-wise) product of two vectors. One can then bound the expected value of  $|f(z)|$  as a univariate polynomial in  $p$ . We then use an extremal property of univariate polynomials (Chebyshev's inequality) to choose  $p$ .

### 3.2. Relating $\Lambda(f)$ to $\|f\|_2$ for Multilinear Polynomials

Here, we use the fact that for a multi-linear polynomial  $f = \sum_{\beta} f_{\beta} x^{\beta}$ ,  $|\mathcal{O}(\beta)| = d!$  for each  $\beta$  with a non-zero coefficient. Note that considering the matrix  $M_f$  with

$$M_f[I, J] = \frac{f_{\alpha(I) + \alpha(J)}}{|\mathcal{O}(\alpha(I) + \alpha(J))|}$$

gives a matrix representation of  $f$ . By the Gershgorin circle theorem, we can bound  $\|M_f\|_2$ , and hence  $\Lambda(f)$ , by  $n^{d/2} \cdot (\max_{\beta} |f_{\beta}| / d!)$ .

On the other hand for a multilinear polynomial, using  $x = \beta / \sqrt{|\beta|}$  (considering  $\beta$  as a vector in  $\mathbb{R}^n$ ), gives  $\|f\|_2 \geq d^{-d/2} \cdot |f_\beta|$  for any  $\beta$ . Thus, we easily get

$$\Lambda(f) \leq \frac{d^{d/2}}{d!} \cdot n^{d/2} \cdot \|f\|_2 = O_d((n/d)^{d/2}) \cdot \|f\|_2.$$

This can now be combined with the reduction from the previous section to obtain an  $(n/q)^{d/2}$  approximation using  $q$  levels of the SoS hierarchy. For an arbitrary  $f$  of degree  $d$ , and for  $q$  divisible by  $2d$ , let  $f^{q/d}(x) = \sum_\alpha F_{2\alpha}(x) \cdot x^\alpha$  be decomposition into multi-linear polynomials  $F_{2\alpha}$ . Then,

$$\frac{\Lambda(f^{q/d})}{\|f\|_2^{q/d}} = \frac{\Lambda(f^{q/d})}{\|f^{q/d}\|_2} \leq 2^{O(q)} \cdot \max_\alpha \frac{\Lambda(F_{2\alpha})}{\|F_{2\alpha}\|_2} \leq 2^{O(q)} \cdot (n/q)^{q/2}.$$

Thus,  $\Lambda(f^{q/d})^{d/q} = O_d((n/q)^{d/2}) \cdot \|f\|_2$ . For the case of polynomials with non-negative coefficients, one can derive better tradeoffs between  $\Lambda(f)$  and  $\|f\|_2$  in the multilinear case, to obtain the improved guarantees.

### 3.3. Improved Approximations via Folding

To obtain the results with additive improvements in the exponent  $((n/q)^{d/2-1})$  in the case of general polynomials and  $(n/q)^{d/4-1/2}$  in the case of polynomials with non-negative coefficients), we need to use the fact that the problem of optimizing quadratic polynomials can be solved in polynomial time. In particular, given a degree- $d$  polynomial  $f$ , we consider its representation as a degree- $(d-2, 2)$  folded polynomial  $h$ , such that  $U(h) = f$ . Let  $h$  be of the form

$$h(x) = \sum_{\beta \in \mathbb{N}_{d-2}^n} \bar{h}_\beta(x) \cdot x^\beta,$$

where each  $\bar{h}_\beta$  is a quadratic polynomial. Recall that (without folding) we could write  $f = \sum_\alpha F_{2\alpha}(x) \cdot x^{2\alpha}$  and argue that

$$\|f\|_2 \geq 2^{-O(d)} \cdot \max_\alpha \|F_{2\alpha}\|_2 \geq 2^{-O(d)} \cdot d^{-d/2} \cdot \max_\alpha \max_\gamma |(F_{2\alpha})_\gamma| = 2^{-O(d)} \cdot d^{-d/2} \cdot \max_\beta |f_\beta|$$

In the case of the folded polynomial, we will instead be able to show

$$\|f\|_2 = \|U(h)\|_2 \geq 2^{-O(d)} \cdot (d-2)^{-(d-2)/2} \cdot \max_{\beta \leq \mathbf{1}} \|\bar{h}_\beta\|_2.$$

Note that the above bound only addresses the folds corresponding to multilinear monomials  $\beta$ . Also, for the case when we consider a power  $f^{q/d}$  of  $f$ , we need to consider degree- $(q-2q/d, 2q/d)$  folded polynomials, since we want to use the fact that any product of  $q/d$  quadratic polynomials is “easy”. We prove a result with general powers and arbitrary monomials in the proof of Theorem 6.8.

For the upper bound on  $\Lambda(f)$ , we again bound it in terms of multilinear components of  $f$ . However, we now need to consider *folded multilinear polynomials*, i.e., we write the folded polynomial  $h$  as

$$h(x) = \sum_{\beta \in \mathbb{N}_{d-2}^n} \bar{h}_\beta(x) \cdot x^\beta = \sum_{|\alpha| \leq (d-2)/2} H_{2\alpha}(x) \cdot x^{2\alpha},$$

where each  $H_{2\alpha}$  is a multilinear polynomial with coefficients in the set  $\{\bar{h}_\beta\}_{\beta \in \mathbb{N}_{d-2}^n}$ . As before, we prove that

$$\Lambda(h) \leq \left(\frac{d}{2} + 1\right) \cdot \max_{\alpha} \frac{\Lambda(H_{2\alpha})}{|\Theta(\alpha)|},$$

where for the folded polynomials  $h$  and  $H_{2\alpha}$ , we use  $\Lambda(\mathcal{U}(h))$  and  $\Lambda(\mathcal{U}(H_{2\alpha}))$  respectively. In fact, we develop a more general version of the above inequality in Lemma 6.2. Finally, to relate  $\Lambda(H_{2\alpha})$  to  $\|f\|_2$ , which is now done by applying a “block” version of the Gershgorin circle theorem (where, instead of sum of absolute values of entries in a row as before, one considers the sum of spectral norms of folds). This is established in Lemma 6.5.

### 3.4. Random Polynomials

#### 3.4.1. Upper Bounds

Let  $A \in \mathbb{R}^{[n]^{d/2} \times [n]^{d/2}}$  be a random matrix whose entry is an independent Rademacher random variable, and let  $f := \sum_{I, J \in [n]^{d/2}} A[I, J] \cdot x^{\alpha(I) + \alpha(J)}$ . It is well known that  $\|f\|_2 \leq O(\sqrt{n} \log d)$  with high probability [TS14]. For such a polynomial  $f$  and an  $q$  divisible by  $d$ , we prove that with high probability,

$$\left(\Lambda(f^{q/d})\right)^{d/q} \leq \tilde{O}_d\left(\left(\frac{n}{q^{1-2/d}}\right)^{d/4}\right) = \tilde{O}_d\left(\left(\frac{n}{q}\right)^{d/4-1/2}\right) \cdot \|f\|_2.$$

**Remark 3.1.** Our techniques prove similar results for a more general random model where each coefficient is independently sampled from a centred subgaussian distribution. See the previous version of the paper [BGL16] for details.

We give an overview of the proof. Let  $d = 4$  for the sake of exposition. To prove an upper bound on  $\Lambda(f)$  using SoS hierarchies of degree  $q$  (assume  $q$  is a multiple of 4), we define a suitable matrix representation  $M := M_{f^{q/4}} \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$  of  $f^{q/4}$  and bounds  $\|M\|_2$ . Since  $\Lambda(f) \leq (\|M\|_2)^{q/4}$  for any representation  $M$ , a good upper bound on  $\|M\|_2$  certifies that  $\Lambda(f)$  is small.

One of the intuitive reasons taking a high power gives a better bound on the spectral norm is that this creates more entries of the matrix that correspond to the same monomial, and distributing the coefficient of this monomial equally among the corresponding entries reduces variance (i.e.,  $\text{Var}[X]$  is less than  $k \cdot \text{Var}[X/k]$  for  $k > 1$ ). In this regard, the most natural representation  $M$  of  $f^{q/4}$  is the *complete symmetrization*.

$$\begin{aligned} M_c[(i_1, \dots, i_{q/2}), (i_{q/2+1}, \dots, i_q)] &= \frac{1}{q!} \cdot \sum_{\pi \in \mathbb{S}_q} A^{\otimes q/4}[(i_{\pi(1)}, \dots, i_{\pi(q/2)}), (i_{\pi(q/2+1)}, \dots, i_{\pi(q)})] \\ &= \frac{1}{q!} \cdot \sum_{\pi \in \mathbb{S}_q} \prod_{j=1}^{q/4} A[(i_{\pi(2j-1)}, i_{\pi(2j)}), (i_{\pi(q/2+2j-1)}, i_{\pi(q/2+2j)})]. \end{aligned}$$

However,  $\|M_c\|_2$  turns out to be much larger than  $\Lambda(f)$ , even when  $q = 8$ . One intuitive explanation is that  $M_c$ , as a  $n^4 \times n^4$  matrix, contains a copy of  $\text{Vec}(A) \text{Vec}(A)^T$ , where  $\text{Vec}(A) \in \mathbb{R}^{n^4}$  is the vector with  $\text{Vec}(A)[i_1, i_2, i_3, i_4] = A[(i_1, i_2), (i_3, i_4)]$ . Then  $\text{Vec}(A)$  is a vector that witnesses

$\|M_c\|_2 \geq \Omega(n^2)$ , regardless of the randomness of  $f$ . Our final representation is the following *row-column independent symmetrization* that simultaneously respects the spectral structure of a random matrix  $A$  and reduces the variance. Our  $M$  is given by

$$\begin{aligned} M[(i_1, \dots, i_{q/2}), (j_1, \dots, j_{q/2})] &= \frac{1}{q!} \cdot \sum_{\pi, \sigma \in S_{q/2}} A^{\otimes q/4}[(i_{\pi(1)}, \dots, i_{\pi(q/2)}), (j_{\sigma(1)}, \dots, j_{\sigma(q/2)})] \\ &= \frac{1}{q!} \cdot \sum_{\pi, \sigma \in S_{q/2}} \prod_{k=1}^{q/4} A[(i_{\pi(2k-1)}, i_{\pi(2k)}), (j_{\sigma(2k-1)}, j_{\sigma(2k)})]. \end{aligned}$$

To formally show  $\|M\|_2 = \tilde{O}(n/\sqrt{q})^{q/4}$  with high probability, we use the trace method to show

$$\mathbb{E} [\text{Tr}(M^p)] \leq 2^{O(pq \log p)} \frac{n^{pq/4+q/2}}{q^{pq/8}},$$

where  $\mathbb{E} [\text{Tr}(M^p)]$  can be written as (let  $I^{p+1} := I^1$ )

$$\mathbb{E} \left[ \sum_{I^1, \dots, I^p \in [n]^{q/2}} \prod_{j=1}^p M[I^j, I^{j+1}] \right] = \sum_{I^1, \dots, I^p} \mathbb{E} \left[ \prod_{j=1}^p \left( \sum_{\pi_j, \sigma_j \in S_{q/2}} \prod_{k=1}^{q/4} A[(I_{\pi_j(2k-1)}^k, I_{\pi_j(2k)}^k), (I_{\sigma_j(2k-1)}^{k+1}, I_{\sigma_j(2k)}^{k+1})] \right) \right].$$

Let  $E(I^1, \dots, I^p)$  be the expectation value for  $I^1, \dots, I^p$  in the right hand side. We study  $E(I^1, \dots, I^p)$  for each  $I^1, \dots, I^p$  by careful counting of the number of permutations on a given sequence with possibly repeated entries. For any  $I^1, \dots, I^p \in [n]^{q/2}$ , let  $\#(I^1, \dots, I^p)$  denote the number of distinct elements of  $[n]$  that occur in  $I^1, \dots, I^p$ , and for each  $s = 1, \dots, \#(I^1, \dots, I^p)$ , let  $c^s \in (\{0\} \cup [q/2])^p$  denote the number of times that the  $j$ th smallest element occurs in  $I^1, \dots, I^p$ . When  $E(I^1, \dots, I^p) \neq 0$ , it means that for some permutations  $\{\pi_j, \sigma_j\}_j$ , every term  $A[\cdot, \cdot]$  must appear even number of times. This implies that the number of distinct elements in  $I^1, \dots, I^p$  is at most half the maximal possible number  $pq/2$ . This lemma proves the intuition via graph theoretic arguments.

**Lemma 3.2.** *If  $E(I^1, \dots, I^p) \neq 0$ ,  $\#(I^1, \dots, I^p) \leq \frac{pq}{4} + \frac{q}{2}$ .*

The number of  $I^1, \dots, I^p$  that corresponds to a sequence  $c^1, \dots, c^s$  is at most  $\frac{n^s}{s!} \cdot \frac{((q/2)!)^p}{\prod_{\ell \in [p]} c_\ell^1 \cdot c_\ell^p}$ . Furthermore, there are at most  $2^{O(pq)} p^{pq/2}$  different choices of  $c^1, \dots, c^s$  that corresponds to some  $I^1, \dots, I^p$ . The following technical lemma bounds  $E(I^1, \dots, I^p)$  by careful counting arguments.

**Lemma 3.3.** *For any  $I^1, \dots, I^p$ ,  $E(I^1, \dots, I^p) \leq 2^{O(pq)} \frac{p^{5pq/8}}{q^{3pq/8}} \prod_{\ell \in [p]} c_\ell^1! \dots c_\ell^s!$ .*

Summing over all  $s$  and multiplying all possibilities,

$$\begin{aligned} \mathbb{E} [\text{Tr}(M^p)] &\leq \sum_{s=1}^{pq/4+q/2} \left( 2^{O(pq)} p^{pq/2} \right) \cdot \left( \frac{n^s}{s!} \cdot ((q/2)!)^p \right) \cdot \left( 2^{O(pq)} \frac{p^{5pq/8}}{q^{3pq/8}} \right) \\ &= \max_s 2^{O(pq \log p)} \cdot n^s \cdot \frac{q^{pq/8}}{s!}. \end{aligned}$$

When  $q \leq n$ , the maximum occurs when  $s = pq/4 + q/2$ , so  $\mathbb{E} [\text{Tr}(M^p)] \leq 2^{O(pq \log p)} \cdot \frac{n^{pq/4+q/2}}{q^{pq/8}}$  as desired.



### 3.4.2. Lower Bounds

Let  $A \in \mathbb{R}^{[n]^{d/2} \times [n]^{d/2}}$  be a random matrix whose entry is an independent Rademacher random variable, and let  $f := \sum_{I, J \in [n]^{d/2}} A[I, J] \cdot x^{\alpha(I) + \alpha(J)}$ . To prove Theorem 1.3, we construct a moment matrix  $M$  that is positive semidefinite, SoS-symmetric,  $\text{Tr}(M) = 1$ , and  $\langle A, M \rangle \geq 2^{-O(d)} \cdot \frac{n^{d/4}}{d^{d/4}}$ . At a high level, our construction is  $M := c_1 A + c_2 W$  for some  $c_1, c_2$ , where  $A$  contains entries of  $A$  only corresponding to the multilinear indices, averaged over all SoS-symmetric positions. This gives a large inner product with  $A$ , SoS-symmetry, and nice spectral properties even though it is not positive semidefinite. The most natural way to make it positive semidefinite is adding a copy of the identity matrix, but this will again break the SoS-symmetry.

Our main technical contribution here is the construction of  $W$  that acts like a *SoS-symmetrized identity*. It has the minimum eigenvalue at least  $\frac{1}{2}$ , while the trace being  $n^{d/2} \cdot 2^{O(d)}$ , so the ratio of the average eigenvalue to the minimum eigenvalue is bounded above by  $2^{O(d)}$ , which allows us to prove a tight lower bound. To the best of our knowledge, no such bound was known for SoS-symmetric matrices except small values of  $d = 3, 4$ .

Given  $I, J \in [n]^{d/2}$ , we let  $W[I, J] := \mathbb{E}[x^{\alpha(I) + \alpha(J)}]$ , where  $x_1, \dots, x_n$  are independently sampled from the *Wigner semicircle distribution*, whose probability density function is the semicircle  $f(x) = \frac{2}{\pi} \sqrt{1 - x^2}$ . Since  $\mathbb{E}[x_1^\ell] = 0$  if  $\ell$  is odd and  $\mathbb{E}[x_1^{2\ell}] = \frac{1}{\ell+1} \binom{2\ell}{\ell}$ , which is the  $\ell$ th Catalan number, each entry of  $W$  is bounded by  $2^{O(d)}$  and  $\text{Tr}(W) \leq n^{d/2} \cdot 2^{O(d)}$ . To prove a lower bound on the minimum eigenvalue, we show that for any degree- $\ell$  polynomial  $p$  with  $m$  variables,  $\mathbb{E}[p(x_1, \dots, x_m)^2]$  is large by induction on  $\ell$  and  $m$ . We use another property of the Wigner semicircle distribution that if  $H \in \mathbb{R}^{(d+1) \times (d+1)}$  is the univariate moment matrix of  $x_1$  defined by  $H[i, j] = \mathbb{E}[x_1^{i+j}]$  ( $0 \leq i, j \leq d$ ) and  $H = (R^T)R$  is the Cholesky decomposition of  $H$ ,  $R$  is an upper triangular matrix with 1's on the main diagonal. This nice Cholesky decomposition allows us to perform the induction on the number of variables while the guarantee on the minimum eigenvalue is independent of  $n$ .

### 3.5. Lower Bounds Non-negative Coefficient Polynomials

We discuss some of the important ideas from the proof of Theorem 1.4. The lower bound given in Theorem 1.3 proves a large ratio  $\frac{\Lambda(f)}{\|f\|_2}$  by considering a random polynomial  $f$  — each coefficient of  $f$  is an independent random variable with bounded variance. Considering such a *completely independent* random instance successfully gave optimal SoS-gap instances for many problems including constraint satisfaction problems [Sch08, Tul09, Cha13, BCK15] and the planted clique problem [MPW15, HKP<sup>+</sup>16a, BHK<sup>+</sup>16].

The most natural adaptation of the above strategy to non-negative degree-4 polynomial is to consider a random polynomial  $f$  where each coefficient  $f_\alpha$  is independently sampled such that  $f_\alpha = 1$  with probability  $p$  and  $f_\alpha = 0$  with probability 0. However, this construction fails for every choice of  $p$ . If we let  $A \in \mathbb{R}^{[n]^2 \times [n]^2}$  be the natural matrix representation of  $f$  (i.e., each coefficient  $f_\alpha$  is distributed uniformly among the corresponding entries of  $A$ ), the Perron-Frobenius theorem shows that  $\|A\|_2$  is less than the maximum row sum  $\max(\tilde{O}(n^2 p), 1)$  of  $M$ , which is also an upper bound on  $\Lambda(f)$ . If we set  $x = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$  (when  $p \geq \frac{1}{n^2}$ ), or take any  $\alpha$  with  $f_\alpha = 1$  and set  $x = \frac{\alpha}{2}$  (when  $p < \frac{1}{n^2}$ ), we can see that  $\|f\|_2 \geq f(x) \geq \max(\Omega(n^2 p), \frac{1}{16})$ , leaving little possible gap between  $\|f\|_2$  and  $\Lambda(f)$ .

We introduce another natural distribution of random non-negative polynomials that bypasses this problem. Let  $G = (V, E)$  be a random graph drawn from the distribution  $G_{n,p}$ , where  $p = \frac{1}{n^{1/3}}$

and  $V = [n]$ . Let  $C_4 \subseteq \binom{V}{4}$  be the set of 4-cliques in  $G$ . The polynomial  $f$  is defined as

$$f(x_1, \dots, x_n) := \sum_{\{i_1, i_2, i_3, i_4\} \in C_4} x_{i_1} x_{i_2} x_{i_3} x_{i_4}.$$

Instead of trying  $\Theta(n^4)$   $p$ -biased random bits, we use  $\Theta(n^2)$  of them. This limited independence bypasses the problem above, since the rows of  $A$  have significantly different row sums.  $\Theta(n^2 p) = \Theta(n^{5/3})$  rows that correspond to an edge of  $G$  will have row sum  $\Theta(n^2 p^5) = \Theta(n^{1/3})$ , and all other rows will be zeros. Since these  $n^2 p$  rows (edges) are chosen independently from  $\binom{[n]}{2}$ , they still reveal little information that can be exploited to find a  $n$ -dimensional vector  $x$  with large  $f(x)$ .

It is simple to prove that  $\|f\|_{sp} \geq \Omega(n^{1/6})$  by considering the Frobenius norm of the  $n^2 p \times n^2 p$  principal submatrix, over any matrix representation (indeed,  $A$  is the representation that minimizes the Frobenius norm). To prove  $\Lambda(f) \geq \tilde{\Omega}(n^{1/6})$ , we construct a moment matrix  $M$  that is SoS-symmetric, positive semidefinite, and has a large  $\langle M, A \rangle$ . It turns out that the  $n^2 p \times n^2 p$  submatrix of  $A$  shares spectral properties of the adjacency matrix of a random graph  $G_{n^2 p, p^4}$ , and taking  $M := c_1 A + c_2 I$  for some identity-like matrix  $I$  proves  $\Lambda(f) \geq \tilde{\Omega}(n^{1/6})$ .

To upper bound  $\|f\|_2$ , we first observe that  $\|f\|_2$  is the same as the following natural combinatorial problem up to an  $O(\log^4 n)$  factor: find four sets  $S_1, S_2, S_3, S_4 \subseteq V$  that maximizes

$$\frac{|C(S_1, S_2, S_3, S_4)|}{\sqrt{|S_1||S_2||S_3||S_4|}}$$

where  $|C(S_1, S_2, S_3, S_4)|$  is the number of 4-cliques  $\{v_1, \dots, v_4\}$  of  $G$  with  $v_i \in S_i$  for  $i = 1, \dots, 4$ . The number of copies of a fixed subgraph  $H$  in  $G_{n,p}$ , including its tail bound, has been actively studied in probabilistic combinatorics [Vu01, KV04, JOR04, Cha12, DK12a, DK12b, LZ16], though we are interested in bounding the 4-clique density of *every* all 4-tuple of subsets simultaneously. The previous results give a strong enough tail bound for union bound to prove that the optimal value of the problem is  $\log^{O(1)} n$  via union bound if we require  $S_1 = \dots = S_4$ , but this strategy inherently does not work when the set sizes become different. We extend previous techniques and introduce additional combinatorial ideas to prove  $\|f\|_2 \leq \tilde{O}(n^{1/12})$ . It is an interesting open question to see whether  $\|f\|_2 = \log^{O(1)} n$ .

## 4. Additional Preliminaries and the SoS Hierarchy

### 4.1. Pseudoexpectations and Moment Matrices

Let  $\mathbb{R}[x]_{\leq q}$  be the vector space of polynomials with real coefficients in variables  $x = (x_1, \dots, x_n)$ , of degree at most  $q$ . For an even integer  $q$ , the degree- $q$  pseudo-expectation operator is a linear operator  $\tilde{\mathbb{E}} : \mathbb{R}[x]_{\leq q} \mapsto \mathbb{R}$  such that

1.  $\tilde{\mathbb{E}}[1] = 1$  for the constant polynomial 1.
2.  $\tilde{\mathbb{E}}[p_1 + p_2] = \tilde{\mathbb{E}}[p_1] + \tilde{\mathbb{E}}[p_2]$  for any polynomials  $p_1, p_2 \in \mathbb{R}[x]_{\leq q}$ .
3.  $\tilde{\mathbb{E}}[p^2] \geq 0$  for any polynomial  $p \in \mathbb{R}[x]_{\leq q/2}$ .

The pseudo-expectation operator  $\tilde{\mathbb{E}}$  can be described by the moment matrix  $\hat{M} \in \mathbb{R}^{\mathbb{N}_{\leq q/2}^n \times \mathbb{N}_{\leq q/2}^n}$  such that  $\hat{M}[\alpha, \beta] = \tilde{\mathbb{E}}[x^{\alpha+\beta}]$  for  $\alpha, \beta \in \mathbb{N}_{\leq q/2}^n$ .

For each fixed  $t \leq q/2$ , we can also consider the principal minor of  $\hat{M}$  indexed by  $\alpha, \beta \in \mathbb{N}_t^n$ . This also defines a matrix  $M \in \mathbb{R}^{[n]^t \times [n]^t}$  with  $M[I, J] = \tilde{\mathbf{E}} \left[ x^{\alpha(I) + \alpha(J)} \right]$ . Note that this new matrix  $M$  satisfies  $M[I, J] = M[K, L]$  whenever  $\alpha(I) + \alpha(J) = \alpha(K) + \alpha(L)$ . Recall that a matrix in  $\mathbb{R}^{[n]^t \times [n]^t}$  with this symmetry is said to be SoS-symmetric.

We will use the following facts about the operator  $\tilde{\mathbf{E}}$  given by the SoS hierarchy.

**Claim 4.1 (Pseudo-Cauchy-Schwarz [BKS14]).**  $\tilde{\mathbf{E}}[p_1 p_2] \leq (\tilde{\mathbf{E}}[p_1^2] \tilde{\mathbf{E}}[p_2^2])^{1/2}$  for any  $p_1, p_2$  of degree at most  $q/2$ .

#### 4.1.1. Constrained Pseudoexpectations

For a system of polynomial constraints  $C = \{f_1 = 0, \dots, f_m = 0, g_1 \geq 0, \dots, g_r \geq 0\}$ , we say  $\tilde{\mathbf{E}}_C$  is a pseudoexpectation operator respecting  $C$ , if in addition to the above conditions, it also satisfies

1.  $\tilde{\mathbf{E}}_C[p \cdot f_i] = 0, \forall i \in [m]$  and  $\forall p$  such that  $\deg(p \cdot f_i) \leq q$ .
2.  $\tilde{\mathbf{E}}_C[p^2 \cdot \prod_{i \in S} g_i] \geq 0, \forall S \subseteq [r]$  and  $\forall p$  such that  $\deg(p^2 \cdot \prod_{i \in S} g_i) \leq q$ .

It is well-known that such constrained pseudoexpectation operators can be described as solutions to semidefinite programs of size  $n^{O(q)}$  [BS14, Lau09]. This hierarchy of semidefinite programs for increasing  $q$  is known as the SoS hierarchy.

#### 4.2. Matrix Representations of Polynomials and $\Lambda(f)$

For a homogeneous polynomial  $f$  of even degree  $d$ , we say a matrix  $M \in \mathbb{R}^{[n]^{d/2} \times [n]^{d/2}}$  is a degree- $d$  matrix representation of  $f$  if for all  $x$ ,  $f(x) = (x^{\otimes d/2})^T \cdot M \cdot x^{\otimes d/2}$ . Recall that we consider the semidefinite program for optimizing the quantity

$$\Lambda(f) = \inf \left\{ \sup_{\|z\|=1} z^T M z \mid M \in \mathbb{S}^{n^{d/2}}(\mathbb{R}), (x^{\otimes d/2})^T \cdot M \cdot x^{\otimes d/2} = f(x) \ \forall x \in \mathbb{R}^n \right\}.$$

It is easy to check that the dual of the above semidefinite program is given by  $\max \tilde{\mathbf{E}}_C[f]$  where  $\tilde{\mathbf{E}}_C$  is a degree- $d$  pseudoexpectation operator respecting the constraint  $C = \{\|x\|_2^d = 1\}$ . Moreover, strong duality holds in this case, since the solution  $\tilde{\mathbf{E}}_C[x^\alpha] = (1/\sqrt{n})^{|\alpha|}$  for all  $\alpha \in \mathbb{N}_{\leq d}^n$  is strictly feasible and in the relative interior of the domain. Thus, the objective values of the two programs are equal.

We will also need to consider constraint sets  $C = \{\|x\|_2^2 = 1, x^{\beta_1} \geq 0, \dots, x^{\beta_m} \geq 0\}$ . We refer to the non-negativity constraints here as moment non-negativity constraints. When considering the maximum of  $\tilde{\mathbf{E}}_C[f]$ , for constraint sets  $C$  containing moments non-negativity constraints in addition to  $\|x\|_2^2 = 1$ , we refer to the optimum value as  $\Lambda_C(f)$ . Note that the maximum is still taken over degree- $d$  pseudoexpectations. Also, strong duality still holds in this case since  $\tilde{\mathbf{E}}_C[x^\alpha] = (1/\sqrt{n})^{|\alpha|}$  is still a strictly feasible solution.

We use the following claim, which is an easy consequence of the fact that the sum-of-squares algorithm can produce a certificate of optimality (see [OZ13]). In particular, if  $\max_{\tilde{\mathbf{E}}_C} \tilde{\mathbf{E}}_C[f] = \Lambda_C(f)$  for a degree- $q_1$  pseudoexpectation operator respecting  $C$  containing  $\|x\|_2^2 = 1$  and moment non-negativity constraints for  $\beta_1, \dots, \beta_m$ , then for every  $\lambda > \Lambda_C(f)$ , we have that  $\lambda - f$  can be

certified to be positive by showing that  $\lambda - f \in \Sigma_C^{q_1}$ . Here  $\Sigma_C^{(q_1)}$  is the set of all expressions of the form

$$\lambda - f = \sum_j p_j \cdot (\|x\|_2^2 - 1) + \sum_{S \subseteq [m]} h_S(x) \cdot \prod_{i \in S} x^{\beta_i},$$

where each  $h_S$  is a sum of squares of polynomials and the degree of each term is at most  $q_1$ .

**Lemma 4.2.** *Let  $\Lambda_C(f)$  denote the maximum of  $\tilde{\mathbf{E}}_C[f]$  over all degree- $d$  pseudoexpectation operators respecting  $C$ . Then, for a pseudoexpectation operator of degree  $d'$  (respecting  $C$ ) and a polynomial  $p$  of degree at most  $(d' - d)/2$ , we have that*

$$\tilde{\mathbf{E}}_C[p^2 \cdot f] \leq \tilde{\mathbf{E}}_C[p^2] \cdot \Lambda_C(f).$$

*Proof.* As described above, for any  $\lambda > \Lambda_C(f)$ , we can write  $\lambda - f = g$  for  $g \in \Sigma_C^{(d)}$ . Since the degree of each term in  $p^2 \cdot g$  is at most  $d'$ , we have by the properties of pseudoexpectation operators (of degree  $d'$ ) that

$$\lambda \cdot \tilde{\mathbf{E}}_C[p^2] - \tilde{\mathbf{E}}_C[p^2 \cdot f] = \tilde{\mathbf{E}}_C[p^2 \cdot (\lambda - f)] = \tilde{\mathbf{E}}_C[p^2 \cdot g] \geq 0. \quad \blacksquare$$

The following monotonicity claim for non-negative coefficient polynomials will come in handy in later sections.

**Lemma 4.3.** *Let  $C$  be a system of polynomial constraints containing the constraints  $\{\forall \beta \in \mathbb{N}_t^n, x^\beta \geq 0\}$ . Then for any non-negative coefficient polynomials  $f$  and  $g$  of degree  $t$ , and such that  $f \geq g$  (coefficient-wise, i.e.  $f - g$  has non-negative coefficients), we have  $\Lambda_C(f) \geq \Lambda_C(g)$ .*

*Proof.* Observe that for any pseudo-expectation operator  $\tilde{\mathbf{E}}_C$  respecting  $C$ , we have  $\tilde{\mathbf{E}}_C[f - g] \geq 0$  because of the moment non-negativity constraints and by linearity.

So let  $\tilde{\mathbf{E}}_C$  be a pseudo-expectation operator realizing  $\Lambda_C(g)$ . Then we have,

$$\Lambda_C(f) \geq \tilde{\mathbf{E}}_C[f] = \tilde{\mathbf{E}}_C[g] + \tilde{\mathbf{E}}_C[f - g] = \Lambda_C(g) + \tilde{\mathbf{E}}_C[f - g] \geq 0. \quad \blacksquare$$

## 5. Results for Polynomials in $\mathbb{R}_d[x]$ and $\mathbb{R}_d^+[x]$

### 5.1. Detailed Overview

We will give an overview for the case of non-negative coefficient (nnc) polynomials, but most of the ideas apply to the general setting as well. We will also refrain from writing out  $2^{O(q)}$  factors.

We would like to analyse how well degree- $q$  SoS approximates the 2-norm of an nnc degree- $d$  homogeneous polynomial. Specifically, we'd like to show that the approximation factor is at most  $(n/q)^{d/4}$ .

#### 5.1.1. Sufficient to analyse SoS-degree = Polynomial-degree regime

We start with the somewhat counter-intuitive observation that nailing down the approximation factor of degree- $d$  SoS (on degree- $d$  homogeneous polynomials) upto  $d^{O(d)}$  factors is sufficient for our end goal. Since we normally imagine  $d$  to be a constant,  $d^{O(d)}$  factors might seem negligible, however note that understanding the regime of SoS-degree = polynomial-degree allows us to apply the result to  $f^{q/d}$ . Indeed, assume that  $\Lambda(g) / \|g\|_2 \leq (n/t)^{t/4}$  for all nnc polynomials  $g$  of degree- $t$ . Then we get,  $\Lambda(f^{q/d})^{d/q} / \|f\|_2 = (\Lambda(f^{q/d}) / \|f^{q/d}\|_2)^{d/q} \leq ((n/q)^{q/4})^{d/q} = (n/q)^{d/4}$ .

### 5.1.2. SoS-degree = Polynomial-degree Regime and Multilinearity

We next note that it is fairly straightforward to save  $d^{O(d)}$  factors when the polynomial is multilinear. Specifically, when SoS-degree = Polynomial degree and the polynomial is multilinear, we observe that the entries of the SoS-symmetric matrix representation of the polynomial are precisely the coefficients of the polynomial scaled down by  $d! \sim d^d$  making it easy to win  $d^{O(d)}$  factors (see Theorem 5.15 and Theorem 5.13 for details).

**Remark.** We note that in order to exploit multilinearity, one need not always consider the SoS-symmetric representation. For instance in the case of a random polynomial  $f$ , it turns out that the SoS-symmetric matrix representation of the induced multilinear polynomials of  $f^{q/d}$  is far from optimal (in terms of spectral norm) and instead, the right representation comes from taking a tensor power of a matrix representation of  $f$  and then symmetrizing rows and columns independently. Nevertheless, one is still exploiting the fact that the size of the orbit of every non-zero coefficient is large when the polynomial is multilinear.

The above approach of taking the SoS-symmetric representation completely breaks down when the polynomial is non-multilinear, which prompts our next step.

### 5.1.3. Decomposing Polynomials into Multilinear components

Observe that the set of monomials with  $\ell$  distinct elements is roughly  $n^\ell$ . So we see that the monomials containing  $\ell$  distinct elements when collected together, behave like a degree- $\ell$  multilinear polynomial as far as number of terms and orbit sizes go; and it turns out that this reflects in the approximation factor as well. When  $\ell$  is smaller than  $d$  the ‘effective degree’ is lower than  $d$ , and so intuitively the approximation factor should have a better exponent in  $n$ .

Though we can’t directly apply the techniques from the multilinear case to non-multilinear polynomials, we will attempt to reduce the problem to a number of multilinear problems wherein we exploit a trade-off between multilinearity and effective-degree. To this end, consider the following decomposition:

**Lemma 5.1.** *Any homogeneous  $n$ -variate degree- $d$  polynomial  $f(x)$  has a unique representation of the form*

$$\sum_{\alpha \in \mathbb{N}_{\leq d/2}^n} F_{2\alpha}(x) \cdot x^{2\alpha}$$

where for any  $\alpha \in \mathbb{N}_{\leq d/2}^n$ ,  $F_{2\alpha}$  is a homogeneous multilinear degree- $(d - 2|\alpha|)$  polynomial.

It turns out that the above decomposition is precisely what we need and the bulk of our technical work will be in establishing (Theorem 5.12)

$$\frac{\Lambda(f)}{\|f\|_2} \leq \max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \frac{\Lambda(F_{2\alpha})}{\|F_{2\alpha}\|_2} \cdot 2^{O(d)} \quad (5.1)$$

which combined with bounds for multilinear polynomials would imply the desired result in both the general and nnc cases.

#### 5.1.4. Reduction to the Multilinear components

The motivation for choosing the aforementioned decomposition (namely peeling off squared monomials) is precisely so that we are able to exhibit a matrix representation of any polynomial  $f$  that is the sum of a few block-diagonal matrices whose blocks are (upto scaling) matrix representations of the polynomials of the form  $F_{2\alpha}$ . In other words, the decomposition is primarily motivated by being able to show that SoS can "split up" the multilinear components without much loss.

Ideally one would like to show Eq. (5.1) by establishing (1):  $\Lambda(f) \leq 2^{O(d)} \cdot \max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \Lambda(F_{2\alpha})$  and (2):  $\|f\|_2 \geq 2^{-O(d)} \cdot \max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \|F_{2\alpha}\|_2$ . However the latter inequality is false and in fact one cannot expect to lose a factor any less than  $|S(\alpha)|^{\Omega(|\alpha|)}$  which unfortunately can be as large as  $d^{\Omega(d)}$  - which is exactly the type of factor we were hoping to gain. This suggests that we must prove a stronger version of inequality (1) - one that is as strong as possible (upto  $2^{O(d)}$  factors), and we'd need to prove the least weak version of inequality (2) as possible.

We strengthen inequality (1) by averaging the "best" representation of  $F_{2\alpha}$  over the  $|\mathcal{O}(\alpha)|$  available diagonal-blocks in the approximately-block-diagonal representation of  $f$  we mentioned earlier. This is the content of Lemma 5.2 wherein we show

$$\Lambda(f) \leq 2^{O(d)} \cdot \max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \frac{\Lambda(F_{2\alpha})}{|\mathcal{O}(\alpha)|}.$$

Quite remarkably, it turns out that the above "strengthened" inequality is "just" enough. Indeed one can show

$$\|f\|_2 \geq 2^{-O(d)} \cdot \max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \frac{\|F_{2\alpha}\|_2}{|\mathcal{O}(\alpha)|} \quad (5.2)$$

which implies Eq. (5.1).

**Non-Negative Coefficient Case.** As a warm-up, we first discuss Eq. (5.2) in the context of nnc polynomials. This is because the nnc assumption implies the optimizer of the 2-norm of  $F_{2\alpha}$  is non-negative, and allows one to circumvent cancellation issues whenever substituting non-negative vectors. We first observe that the term  $x^{2\alpha}$  multiplies  $F_{2\alpha}$  in the expansion of  $f$ . This suggests that if one wanted to exhibit a unit vector  $y$  such that  $f(y)$  is large, one should start with the optimizer of  $F_{2\alpha}$  and place extra mass on the coordinates in  $S(\alpha)$ . However, this increases the length of the vector as a function of  $\alpha$ , and one needs to scale down by length of the vector which tells us that the distribution of mass is crucial. It turns out that in order to lose only a factor of  $|\mathcal{O}(\alpha)|$ , the right way to reweight the coordinates of the optimizer of  $F_{2\alpha}$  is to set the coordinates proportional to  $\sqrt{\alpha_i}$ .

**General Case.** For the general version of Eq. (5.2), the reweighting is similar to the non-negative case, however cancellations become the primary issue. To fix this, we resort to a delicate first moment method and this is the most technical section (Lemma 5.11) in the proof of Theorem 5.12. To do this however, we required the use of complex random variables and vectors, and so we first establish that  $\|f\|_2^c \geq \|f\|_2 \cdot 2^{-O(d)}$ , where  $\|f\|_2^c$  is the max of  $|f(z)|$  over all  $z$  lying on the complex unit sphere. The proof of this fact (Lemma 5.8) involves an averaging argument followed by an application of the so-called polarization (decoupling) lemma - Lemma 5.6.



## 5.2. Reduction to Multilinear Polynomials

We would like to approximate  $\|f\|_2$  by individually approximating  $\|F_{2\alpha}\|_2$  for each multilinear polynomial  $F_{2\alpha}$ . This section will establish the soundness of this goal.

### 5.2.1. Upper Bounding $\Lambda(f)$ in terms of $\Lambda(F_{2\alpha})$

We first bound  $\Lambda(f)$  in terms of  $\max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \Lambda(F_{2\alpha})$ . The basic intuition is that any matrix  $M_f$  such that  $(x^{\otimes(d/2)})^T \cdot M_f \cdot x^{\otimes(d/2)}$  for all  $x$  (called a matrix representation of  $f$ ) can be written as a sum of matrices  $M_{t,f}$  for each  $t \leq d/2$ , each of which is block-diagonal matrix with blocks corresponding to matrix representations of the polynomials  $M_{F_{2\alpha}}$  for each  $\alpha$  with  $|\alpha| = 2t$ .

**Lemma 5.2.** *Consider any homogeneous  $n$ -variate degree- $d$  polynomial  $f(x)$ . We have,*

$$\Lambda(f) \leq \max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \frac{\Lambda(F_{2\alpha})}{|\mathcal{O}(\alpha)|} (1 + d/2)$$

*Proof.* We shall start by constructing an appropriate matrix representation  $M_f$  of  $f$  that will give us the desired upper bound on  $\Lambda(f)$ . To this end, for any  $\alpha \in \mathbb{N}_{\leq d/2}^n$ , let  $M_{F_{2\alpha}}$  be the matrix representation of  $F_{2\alpha}$  realizing  $\Lambda(F_{2\alpha})$ . For any  $0 \leq t \leq d/2$ , we define  $M_{(t,f)}$  so that for any  $\alpha \in \mathbb{N}_t^n$  and  $I \in \mathcal{O}(\alpha)$ ,  $M_{(t,f)}[I, I] := M_{F_{2\alpha}} / |\mathcal{O}(\alpha)|$ , and  $M_{(t,f)}$  is zero everywhere else. Now let  $M_f := \sum_{t \in [d/2]} M_{(t,f)}$ . As for validity of  $M_f$  as a representation of  $f$  we have,

$$\begin{aligned} \langle M_f, x^{\otimes d/2} (x^{\otimes d/2})^T \rangle &= \sum_{0 \leq t \leq \frac{d}{2}} \langle M_{(t,f)}, x^{\otimes d/2} (x^{\otimes d/2})^T \rangle \\ &= \sum_{\alpha \in \mathbb{N}_{\leq d/2}^n} \sum_{I \in \mathcal{O}(\alpha)} \langle M_{(|\alpha|,f)}[I, I], x^{\otimes(d/2-|\alpha|)} (x^{\otimes(d/2-|\alpha|)})^T \rangle x^{2\alpha} \\ &= \sum_{\alpha \in \mathbb{N}_{\leq d/2}^n} \sum_{I \in \mathcal{O}(\alpha)} \frac{1}{|\mathcal{O}(\alpha)|} \langle M_{F_{2\alpha}}, x^{\otimes(d/2-|\alpha|)} (x^{\otimes(d/2-|\alpha|)})^T \rangle x^{2\alpha} \\ &= \sum_{\alpha \in \mathbb{N}_{\leq d/2}^n} x^{2\alpha} \cdot \langle M_{F_{2\alpha}}, x^{\otimes(d/2-|\alpha|)} (x^{\otimes(d/2-|\alpha|)})^T \rangle \\ &= \sum_{\alpha \in \mathbb{N}_{\leq d/2}^n} F_{2\alpha}(x) x^{2\alpha} \\ &= f(x) \end{aligned}$$

Now observe that  $M_{(t,f)}$  is a block-diagonal matrix (up to simultaneous permutation of its rows and columns). Thus we have  $\|M_{(t,f)}\| \leq \max_{\alpha \in \mathbb{N}_t^n} \|M_{F_{2\alpha}}\| / |\mathcal{O}(\alpha)|$ . Thus on applying triangle inequality, we obtain  $\|M_f\| \leq \max_{\alpha \in \mathbb{N}_{\leq d/2}^n} (1 + d/2) \|M_{F_{2\alpha}}\| / |\mathcal{O}(\alpha)|$  ■

### 5.2.2. Lower Bounding $\|f\|_2$ in terms of $\|F_{2\alpha}\|_2$ (non-negative coefficients)

We first bound  $\|f\|_2$  in terms of  $\max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \|F_{2\alpha}\|_2$ , when every coefficient of  $f$  is non-negative. If  $x^*$  is the optimizer of  $F_{2\alpha}$ , then it is easy to see that  $x^* \geq 0$ . Setting  $y = x^* + \frac{\sqrt{\alpha}}{|\alpha|}$  ensures that  $\|y\|_2 \leq 2$  and  $f(y)$  is large, since  $f(y)$  recovers a significant fraction (up to a  $2^{O(d)} \cdot |\mathcal{O}(\alpha)|$  factor) of  $F_{2\alpha}(x^*)$ .

**Lemma 5.3.** Let  $f(x)$  be a homogeneous  $n$ -variate degree- $d$  polynomial with non-negative coefficients. Consider any  $\alpha \in \mathbb{N}_{\leq d/2}^n$ . Then

$$\|f\|_2 \geq \frac{\|F_{2\alpha}\|_2}{2^{O(d)} |\mathcal{O}(\alpha)|}.$$

*Proof.* Consider any  $0 \leq t \leq d/2$ , and any  $\alpha \in \mathbb{N}_t^n$ . Let  $x_\alpha^* := \operatorname{argmax} \|F_{2\alpha}\|_2$  (note  $x_\alpha^*$  must be non-negative). Let

$$y^* := x_\alpha^* + \frac{\sqrt{\alpha}}{\sqrt{t}}$$

and let  $x^* := y^* / \|y^*\|$ . The second term is a unit vector since  $\|\sqrt{\alpha}\|_2^2 = t$ . Thus  $\|y^*\| = \Theta(1)$  since  $y^*$  is the sum of two unit vectors. This implies  $f(x^*) \geq f(y^*) / 2^{O(d)}$ . Now we have,

$$\begin{aligned} f(y^*) &= \sum_{\beta \in \mathbb{N}_{\leq d/2}^n} F_{2\beta}(y^*) \cdot (y^*)^{2\beta} && \text{(by Lemma 5.1)} \\ &\geq F_{2\alpha}(y^*) \cdot (y^*)^{2\alpha} && \text{(by non-negativity of coefficients)} \\ &\geq F_{2\alpha}(y^*) \frac{1}{t^t} \prod_{\ell \in \mathcal{S}(\alpha)} \alpha_\ell^{\alpha_\ell} && (y^* \geq \frac{\sqrt{\alpha}}{\sqrt{t}} \text{ entry-wise}) \\ &\geq F_{2\alpha}(y^*) \frac{1}{2^{O(t)} t!} \prod_{\ell \in \mathcal{S}(\alpha)} \alpha_\ell^{\alpha_\ell} \\ &\geq F_{2\alpha}(y^*) \frac{\prod_{\ell \in \mathcal{S}(\alpha)} \alpha_\ell!}{2^{O(t)} t!} \\ &\geq F_{2\alpha}(y^*) \frac{1}{2^{O(t)} |\mathcal{O}(\alpha)|} \\ &\geq F_{2\alpha}(x^*) \frac{1}{2^{O(t)} |\mathcal{O}(\alpha)|} && (y^* \text{ is entry-wise at least } x^*) \\ &= \frac{\|F_{2\alpha}\|_2}{2^{O(t)} |\mathcal{O}(\alpha)|}. \end{aligned}$$

This completes the proof. ■

**Theorem 5.4.** Consider any homogeneous  $n$ -variate degree- $d$  polynomial  $f(x)$  with non-negative coefficients. Then

$$\frac{\Lambda(f)}{\|f\|_2} \leq 2^{O(d)} \max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \frac{\Lambda(F_{2\alpha})}{\|F_{2\alpha}\|_2}.$$

*Proof.* Combining Lemma 5.2 and Lemma 5.3 yields the claim. ■

We will next generalize Theorem 5.4 by proving a more general version of Lemma 5.3.

### 5.2.3. Lower Bounding $\|f\|_2$ in terms of $\|F_{2\alpha}\|_2$ (general case)

We lower bound  $\|f\|_2$  in terms of  $\|F_{2\alpha}\|_2$  for all polynomials. We will first recollect and establish some polynomial identities that will be used in the proof of the generalized version of Lemma 5.3 (i.e. Lemma 5.11).

## Polynomial Identities

**Lemma 5.5 (Chebyshev's Inequality).** Let  $p(x)$  be a univariate degree- $d$  polynomial and let  $c_d$  be it's leading coefficient. Then we have,  $\max_{x \in [0,1]} |p(x)| \geq 2|c_d|/4^d$ .

**Lemma 5.6 ([HLZ10]).** Let  $x^1, x^2, \dots, x^d \in \mathbb{R}^n$  be arbitrary, let  $\mathcal{A} \in \mathbb{R}^{[n]^d}$  be a SoS-symmetric  $d$ -tensor, and let  $\xi_1, \dots, \xi_d$  be independent Rademacher random variables. Then

$$\mathbb{E} \left[ \prod_{i \in [d]} \xi_i \langle \mathcal{A}, (\xi_1 x^1 + \dots + \xi_d x^d)^{\otimes d} \rangle \right] = d! \langle \mathcal{A}, x^1 \otimes \dots \otimes x^d \rangle.$$

This lemma implies:

**Lemma 5.7 ([HLZ10]).** Let  $\mathcal{A}$  be a SoS-symmetric  $d$ -tensor and let  $f(x) := \langle \mathcal{A}, x^{\otimes d} \rangle$ . Then

$$\|f\|_2 \geq \frac{1}{2^{O(d)}} \max_{\|x_i\|=1} \langle \mathcal{A}, x^1 \otimes \dots \otimes x^d \rangle.$$

**Lemma 5.8.** Let  $f$  be an  $n$ -variate degree- $d$  homogeneous polynomial. Let  $\|f\|_2^\zeta := \max_{\substack{z \in \mathbb{C}^n \\ \|z\|=1}} |f(z)|$ , then

$$\frac{\|f\|_2^\zeta}{2^{O(d)}} \leq \|f\|_2 \leq \|f\|_2^\zeta.$$

*Proof.* Let  $\mathcal{A}$  be the SoS-symmetric tensor representing  $f$ . Let  $z^* = a^* + ib^*$  be the complex unit vector realizing  $f(z^*) = \|f\|_2^\zeta$ . Then we have,

$$\begin{aligned} f(z^*) &= \langle \mathcal{A}, (z^*)^{\otimes d} \rangle \\ &= \langle \mathcal{A}, (a^* + ib^*)^{\otimes d} \rangle \\ &= \sum_{\substack{c^1, \dots, c^d \in \{a^*, ib^*\} \\ j \in [d]}} \langle \mathcal{A}, \bigotimes_{j \in [d]} c^j \rangle \\ \Rightarrow \operatorname{Re}(f(z^*)) &= \sum_{\substack{c^1, \dots, c^d \in \{a^*, b^*\} \\ |\{j | c^j = b^*\}| \% 4 = 0}} \langle \mathcal{A}, \bigotimes_{j \in [d]} c^j \rangle - \sum_{\substack{c^1, \dots, c^d \in \{a^*, b^*\} \\ |\{j | c^j = b^*\}| \% 4 = 2}} \langle \mathcal{A}, \bigotimes_{j \in [d]} c^j \rangle, \\ \operatorname{Im}(f(z^*)) &= \sum_{\substack{c^1, \dots, c^d \in \{a^*, b^*\} \\ |\{j | c^j = b^*\}| \% 4 = 1}} \langle \mathcal{A}, \bigotimes_{j \in [d]} c^j \rangle - \sum_{\substack{c^1, \dots, c^d \in \{a^*, b^*\} \\ |\{j | c^j = b^*\}| \% 4 = 3}} \langle \mathcal{A}, \bigotimes_{j \in [d]} c^j \rangle \end{aligned}$$

which implies that there exists  $c^1, \dots, c^d \in \{a^*, b^*\}$  such that  $|\langle \mathcal{A}, \bigotimes_{j \in [d]} c^j \rangle| \geq \|f\|_2^\zeta / 2^{O(d)}$ . Lastly, applying Lemma 5.7 implies the claim.  $\blacksquare$

## Some Probability Facts

**Lemma 5.9.** Let  $X_1, \dots, X_k$  be i.i.d. Bernoulli( $p$ ) random variables. Then for any  $t_1, \dots, t_k \in \mathbb{N}$ ,  $\mathbb{E} [X_1^{t_1} \dots X_k^{t_k}] = p^k$ .

**Lemma 5.10.** Let  $\zeta$  be a uniformly random  $p$ -th root of unity. Then for any  $t \in [p-1]$ ,  $\mathbb{E} [\zeta^t] = 0$ . Also, clearly  $\mathbb{E} [\zeta^p] = 1$ .

We finally lower bound  $\|f\|_2$  in terms of  $F_{2\alpha}$ . Fix  $\alpha \in \mathbb{N}_{\leq d/2}^n$  and, let  $x^*$  be the optimizer of  $F_{2\alpha}$ . Setting  $y = x^* + \frac{\sqrt{\alpha}}{|\alpha|}$  as in the non-negative coefficient case does not work since terms from  $F_{2\beta}$  may be negative. We bypass this issue by first lower bounding  $\|f\|_2^c$  in terms of  $F_{2\alpha}$  and using Lemma 5.8. For  $\|f\|_2^c$ , we use random roots of unity and Bernoulli random variables, together with Lemma 5.5, to extract nonzero contribution only from the monomials that are multiples of  $x^\alpha$  times multilinear parts.

**Lemma 5.11.** *Let  $f(x)$  be a homogeneous  $n$ -variate degree- $d$  polynomial. Then for any  $\alpha \in \mathbb{N}_{\leq d/2}^n$ ,*

$$\|f\|_2 \geq \frac{\|F_{2\alpha}\|_2}{2^{O(d)} |\mathcal{O}(\alpha)|}.$$

*Proof.* Fix any  $\alpha \in \mathbb{N}_{\leq d/2}^n$ , let  $t := |\alpha|$  and let  $k := d - 2t$ . For any  $i \in [n]$ , let  $\zeta_i$  be an independent and uniformly randomly chosen  $(2\alpha_i + 1)$ -th root of unity, and let  $\Xi$  be an independent and uniformly randomly chosen  $(k + 1)$ -th root of unity.

Let  $\bar{x} := \operatorname{argmax} \|F_{2\alpha}\|_2$ . Let  $p \in [0, 1]$  be a parameter to be fixed later, let  $b_1, \dots, b_n$  be i.i.d. Bernoulli( $p$ ) random variables, let  $\zeta := (\zeta_1, \dots, \zeta_n)$ ,  $b := (b_1, \dots, b_n)$  and finally let

$$z := \Xi \cdot b \circ \frac{1}{2\alpha + \mathbb{1}} \circ \bar{x} + \frac{\zeta \circ \sqrt{\alpha}}{\sqrt{t}}.$$

Since  $\sum_{\ell \in S(\alpha)} \alpha_\ell = t$  and roots of unity have magnitude one,  $z$  has length  $O(1)$ . Now consider any fixed  $\gamma \in \{0, 1\}_k^n$ . We have,

$$\begin{aligned} & \mathbb{E} \left[ z^{2\alpha + \gamma} \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] \\ &= \text{coefficient of } \Xi^k \cdot \prod_{i \in [n]} \zeta_i^{2\alpha_i} \text{ in } \mathbb{E} [z^{2\alpha + \gamma}] \quad (\text{by Lemma 5.10}) \\ &= \text{coefficient of } \Xi^k \cdot \prod_{i \in [n]} \zeta_i^{2\alpha_i} \text{ in } \mathbb{E} \left[ \prod_{i \in [n]} \left( \zeta_i \cdot \frac{\sqrt{\alpha_i}}{\sqrt{t}} + \Xi \cdot \frac{b_i \cdot \bar{x}_i}{2\alpha_i + 1} \right)^{2\alpha_i + \gamma_i} \right] \\ &= \prod_{i \in [n]} \text{coefficient of } \Xi^{\gamma_i} \cdot \zeta_i^{2\alpha_i} \text{ in } \mathbb{E} \left[ \left( \zeta_i \cdot \frac{\sqrt{\alpha_i}}{\sqrt{t}} + \Xi \cdot \frac{b_i \cdot \bar{x}_i}{2\alpha_i + 1} \right)^{2\alpha_i + \gamma_i} \right] \quad (\text{since } \gamma \in \{0, 1\}_k^n) \\ &= p^k \cdot \prod_{i \in S(\alpha)} \frac{\alpha_i^{\alpha_i}}{t^{\alpha_i}} \cdot \bar{x}_i^{\gamma_i} \quad (\text{by Lemma 5.9}) \\ &= p^k \cdot \bar{x}^\gamma \cdot \prod_{i \in S(\alpha)} \frac{\alpha_i^{\alpha_i}}{t^{\alpha_i}} \end{aligned}$$

Thus we have,

$$\begin{aligned} & \mathbb{E} \left[ f(z) \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] \\ &= \sum_{\beta \in \mathbb{N}_d^n} f_\beta \cdot \mathbb{E} \left[ z^\beta \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] \end{aligned}$$

$$= \sum_{\substack{\beta \in \mathbb{N}_d^n \\ \beta \geq 2\alpha}} f_\beta \cdot \mathbb{E} \left[ z^\beta \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] \quad (\text{by Lemma 5.10})$$

$$= \sum_{\gamma \in \{0,1\}_k^n} f_{2\alpha+\gamma} \cdot \mathbb{E} \left[ z^{2\alpha+\gamma} \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] + \sum_{\substack{\gamma \in \mathbb{N}_k^n \\ \gamma \not\leq \mathbf{1}}} f_{2\alpha+\gamma} \cdot \mathbb{E} \left[ z^{2\alpha+\gamma} \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right]$$

$$= \sum_{\gamma \in \{0,1\}_k^n} f_{2\alpha+\gamma} \cdot \mathbb{E} \left[ z^{2\alpha+\gamma} \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] + r(p) \quad (\text{by Lemma 5.9})$$

where  $r(p)$  is some univariate polynomial in  $p$ , s.t.  $\deg(r) < k$

$$= \sum_{\gamma \in \{0,1\}_k^n} f_{2\alpha+\gamma} \cdot p^k \cdot \bar{x}^\gamma \cdot \prod_{i \in S(\alpha)} \frac{\alpha_i^{\alpha_i}}{t^{\alpha_i}} + r(p)$$

$$= p^k \cdot F_{2\alpha}(\bar{x}) \cdot \prod_{i \in S(\alpha)} \frac{\alpha_i^{\alpha_i}}{t^{\alpha_i}} + r(p) \quad (\text{where } \deg(r) < k)$$

Lastly we have,

$$\begin{aligned} \|f\|_2 &\geq \|f\|_2^c \cdot 2^{-O(d)} && \text{by Lemma 5.8} \\ &\geq \max_{p \in [0,1]} \mathbb{E} [|f(z)|] \cdot 2^{-O(d)} && (\|z\| = O(1)) \\ &= \max_{p \in [0,1]} \mathbb{E} \left[ \left| f(z) \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right| \right] \cdot 2^{-O(d)} \\ &\geq \max_{p \in [0,1]} \left| \mathbb{E} \left[ f(z) \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] \right| \cdot 2^{-O(d)} \\ &\geq |F_{2\alpha}(\bar{x})| \cdot \prod_{i \in S(\alpha)} \frac{\alpha_i^{\alpha_i}}{t^{\alpha_i}} \cdot 2^{-O(d)} && (\text{by Chebyshev: Lemma 5.5}) \\ &= \|F_{2\alpha}\|_2 \cdot \prod_{i \in S(\alpha)} \frac{\alpha_i^{\alpha_i}}{t^{\alpha_i}} \cdot 2^{-O(d)} \\ &\geq \frac{\|F_{2\alpha}\|_2}{|\mathcal{O}(\alpha)|} \cdot 2^{-O(d)} \end{aligned}$$

This completes the proof. ■

**Theorem 5.12.** Let  $f(x)$  be a homogeneous  $n$ -variate degree- $d$  (for even  $d$ ) polynomial. Then

$$\frac{\Lambda(f)}{\|f\|_2} \leq 2^{O(d)} \max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \frac{\Lambda(F_{2\alpha})}{\|F_{2\alpha}\|_2}.$$

*Proof.* Combining Lemma 5.2 and Lemma 5.11 yields the claim. ■

### 5.3. $(n/q)^{d/4}$ -Approximation for Non-negative Coefficient Polynomials

**Theorem 5.13.** Consider any homogeneous multilinear  $n$ -variate degree- $d$  polynomial  $f(x)$  with non-negative coefficients. We have,

$$\frac{\Lambda(f)}{\|f\|_2} \leq 2^{O(d)} \frac{n^{d/4}}{d^{d/4}}.$$

*Proof.* Let  $M_f$  be the SoS-symmetric matrix representation of  $f$ . Let  $I^* = (i_1, \dots, i_{d/2}) \in [n]^{d/2}$  be the multi-index of any row of  $M_f$  with maximum row sum. Let  $S_I$  for  $I \in [n]^{d/2}$ , denote the sum of the row  $I$  of  $M_f$ . By Perron-Frobenius theorem,  $\|M_f\| \leq S_{I^*}$ . Thus  $\Lambda(f) \leq S_{I^*}$ .

We next proceed to bound  $\|f\|_2$  from below. To this end, let  $x^* := y^* / \|y^*\|$  where,

$$y^* := \frac{\mathbb{1}}{\sqrt{n}} + \frac{1}{\sqrt{d/2}} \sum_{i \in I^*} e_i$$

Since  $f$  is multilinear,  $I^*$  has all distinct elements, and so the second term in the definition of  $y^*$  is of unit length. Thus  $\|y^*\| = \Theta(1)$ , which implies that  $\|f\|_2 \geq f(x^*) \geq f(y^*)/2^{O(d)}$ . Now we have,

$$\begin{aligned} f(y^*) &= ((y^*)^{\otimes d/2})^T M_f (y^*)^{\otimes d/2} \\ &\geq \sum_{I \in \mathcal{O}(I^*)} \frac{1}{(nd)^{d/4}} e_{I(1)}^T \otimes \dots \otimes e_{I(d/2)}^T M_f \mathbb{1}^{\otimes d/2} \quad (\text{by non-negativity of } M_f) \\ &= \sum_{I \in \mathcal{O}(I^*)} \frac{1}{(nd)^{d/4}} e_I^T M_f \mathbb{1} \quad (\in \mathbb{R}^{[n]^{d/2}}) \\ &= \sum_{I \in \mathcal{O}(I^*)} \frac{S_I}{(nd)^{d/4}} \\ &= \sum_{I \in \mathcal{O}(I^*)} \frac{S_{I^*}}{(nd)^{d/4}} \quad (\text{by SoS-symmetry of } M_f) \\ &= \frac{(d/2)! S_{I^*}}{(nd)^{d/4}} \quad (|\mathcal{O}(I^*)| = (d/2)! \text{ by multilinearity of } f) \\ &\geq \frac{d^{d/4} S_{I^*}}{n^{d/4} 2^{O(d)}} \geq \frac{d^{d/4} \Lambda(f)}{n^{d/4} 2^{O(d)}}. \end{aligned}$$

This completes the proof. ■

**Theorem 5.14.** *Let  $f(x)$  be a homogeneous  $n$ -variate degree- $d$  polynomial with non-negative coefficients. Then for any even  $q$  such that  $d$  divides  $q$ ,*

$$\frac{(\Lambda(f^{q/d}))^{d/q}}{\|f\|_2} \leq 2^{O(d)} \frac{n^{d/4}}{q^{d/4}}.$$

*Proof.* Applying Theorem 5.4 to  $f^{q/d}$  and combining this with Theorem 5.13 yields the claim. ■

#### 5.4. $(n/q)^{d/2}$ Approximation for General Polynomials

**Theorem 5.15.** *Consider any homogeneous multilinear  $n$ -variate degree- $d$  polynomial  $f(x)$ . We have,*

$$\frac{\Lambda(f)}{\|f\|_2} \leq 2^{O(d)} \frac{n^{d/2}}{d^{d/2}}.$$

*Proof.* Let  $M_f$  be a matrix representation of  $f$  given by

$$M_f[I, J] = \frac{f_{\alpha(I) + \alpha(J)}}{|\mathcal{O}(\alpha(I) + \alpha(J))|}.$$



By the Gershgorin circle theorem, we can bound  $\|M_f\|_2$ , and hence  $\Lambda(f)$  by  $n^{d/2} \cdot (\max_\beta |f_\beta|/d!)$ . Here, we use the multilinearity of  $f$ . On the other hand for a multilinear polynomial, using  $x = \beta/\sqrt{|\beta|}$  (where  $|\beta| = d$  by multilinearity), gives  $\|f\|_2 \geq d^{-d/2} \cdot |f_\beta|$ . Thus, we easily get

$$\Lambda(f) \leq \frac{d^{d/2}}{d!} \cdot n^{d/2} \cdot \|f\|_2 = 2^{O(d)} \frac{n^{d/2}}{d^{d/2}}. \quad \blacksquare$$

**Theorem 5.16.** *Let  $f(x)$  be a homogeneous  $n$ -variate degree- $d$  polynomial, and assume that  $2d$  divides  $q$ ,*

$$\frac{(\Lambda(f^{q/d}))^{d/q}}{\|f\|_2} \leq 2^{O(d)} \frac{n^{d/2}}{q^{d/2}}.$$

*Proof.* Applying Theorem 5.12 to  $f^{q/d}$  and combining this with Theorem 5.15 yields the claim.  $\blacksquare$

## 6. Approximating 2-norms via Folding

### 6.1. Preliminaries

Recall that we call a folded polynomial multilinear if all its monomials are multilinear. In particular, there's no restriction on the folds of the polynomial.

**Lemma 6.1 (Folded Analogue of Lemma 5.1).**

*Let  $(\mathbb{R}_{d_2}[x])_{d_1}[x] \ni f(x) := \sum_{\beta \in \mathbb{N}_{d_1}^n} \bar{f}_\beta(x) \cdot x^\beta$  be a  $(d_1, d_2)$ -folded polynomial.  $f$  can be written as*

$$\sum_{\alpha \in \mathbb{N}_{\leq d_1/2}^n} F_{2\alpha}(x) \cdot x^{2\alpha}$$

*where for any  $\alpha \in \mathbb{N}_{\leq d_1/2}^n$ ,  $F_{2\alpha}(x)$  is a multilinear  $(d_1 - 2|\alpha|, d_2)$ -folded polynomial.*

*Proof.* Simply consider the folded polynomial

$$F_{2\alpha}(x) = \sum_{\gamma \in \{0,1\}_{d_1-2|\alpha|}^n} \overline{(F_{2\alpha})_\gamma} \cdot x^\gamma$$

where  $\overline{(F_{2\alpha})_\gamma} = \bar{f}_{2\alpha+\gamma}$ .  $\blacksquare$

### 6.2. Reduction to Multilinear Folded Polynomials

Here we will prove a generalized version of Lemma 5.2, which is a generalization in two ways; firstly it allows for folds instead of just coefficients, and secondly it allows a more general set of constraints than just the hypersphere since we will need to add some additional non-negativity constraints for the case of non-negative coefficient polynomials (so that  $\Lambda_C()$  satisfies monotonicity over nnc polynomials which will come in handy later).

Recall that  $\Lambda_C()$  is defined in Section 4.2 and that  $\|f\|_2$  and  $\Lambda_C(f)$  for a folded polynomial  $f$ , are applied to the unfolding of  $f$ .

### 6.2.1. Relating $\Lambda_C(f)$ to $\Lambda_C(F_{2\alpha})$

**Lemma 6.2 (Folded Analogue of Lemma 5.2).**

Let  $C$  be a system of polynomial constraints of the form  $\{\|x\|_2^2 = 1\} \cup C'$  where  $C'$  is a moment non-negativity constraint set. Let  $f \in (\mathbb{R}_{d_2}[x])_{d_1}[x]$  be a  $(d_1, d_2)$ -folded polynomial. We have,

$$\Lambda_C(f) \leq \max_{\alpha \in \mathbb{N}_{\leq d_1/2}^n} \frac{\Lambda_C(F_{2\alpha})}{|\mathcal{O}(\alpha)|} (1 + d_1/2)$$

*Proof.* Consider any degree- $(d_1 + d_2)$  pseudo-expectation operator  $\tilde{\mathbf{E}}_C$ . We have,

$$\begin{aligned} \tilde{\mathbf{E}}_C[f] &= \sum_{\alpha \in \mathbb{N}_{\leq d_1/2}^n} \tilde{\mathbf{E}}_C[F_{2\alpha}(x) \cdot x^{2\alpha}] && \text{(by Lemma 6.1)} \\ &\leq \sum_{\alpha \in \mathbb{N}_{\leq d_1/2}^n} \tilde{\mathbf{E}}_C[x^{2\alpha}] \cdot \Lambda_C(F_{2\alpha}) && \text{(by Lemma 4.2)} \\ &= \sum_{0 \leq t \leq \frac{d_1}{2}} \sum_{\alpha \in \mathbb{N}_t^n} \tilde{\mathbf{E}}_C[x^{2\alpha}] \cdot \Lambda_C(F_{2\alpha}) \\ &= \sum_{0 \leq t \leq \frac{d_1}{2}} \sum_{\alpha \in \mathbb{N}_t^n} \tilde{\mathbf{E}}_C[|\mathcal{O}(\alpha)| x^{2\alpha}] \cdot \frac{\Lambda_C(F_{2\alpha})}{|\mathcal{O}(\alpha)|} \\ &\leq \sum_{0 \leq t \leq \frac{d_1}{2}} \sum_{\alpha \in \mathbb{N}_t^n} \tilde{\mathbf{E}}_C[|\mathcal{O}(\alpha)| x^{2\alpha}] \cdot \max_{\beta \in \mathbb{N}_{\leq d_1/2}^n} \frac{\Lambda_C(F_{2\beta})}{|\mathcal{O}(\beta)|} && (\tilde{\mathbf{E}}_C[x^{2\alpha}] \geq 0) \\ &= \sum_{0 \leq t \leq \frac{d_1}{2}} \tilde{\mathbf{E}}_C \left[ \sum_{\alpha \in \mathbb{N}_t^n} |\mathcal{O}(\alpha)| x^{2\alpha} \right] \cdot \max_{\beta \in \mathbb{N}_{\leq d_1/2}^n} \frac{\Lambda_C(F_{2\beta})}{|\mathcal{O}(\beta)|} \\ &= \sum_{0 \leq t \leq \frac{d_1}{2}} \tilde{\mathbf{E}}_C[\|x\|_2^{2t}] \cdot \max_{\beta \in \mathbb{N}_{\leq d_1/2}^n} \frac{\Lambda_C(F_{2\beta})}{|\mathcal{O}(\beta)|} \\ &= \sum_{0 \leq t \leq \frac{d_1}{2}} \max_{\beta \in \mathbb{N}_{\leq d_1/2}^n} \frac{\Lambda_C(F_{2\beta})}{|\mathcal{O}(\beta)|} \\ &= \max_{\beta \in \mathbb{N}_{\leq d_1/2}^n} \frac{\Lambda_C(F_{2\beta})}{|\mathcal{O}(\beta)|} (1 + d_1/2) \end{aligned}$$

■

### 6.3. Relating Evaluations of $f$ to Evaluations of $F_{2\alpha}$

Here we would like to generalize Lemma 5.3 and Lemma 5.11 to allow folds, however for technical reasons related to decoupling of the domain of the folds from the domain of the monomials of a folded polynomial, we instead generalize claims implicit in the proofs of Lemma 5.3 and Lemma 5.11.

Let  $f \in (\mathbb{R}_{d_2}[x])_{d_1}[x]$  be a  $(d_1, d_2)$ -folded polynomial. Recall that an evaluation of a folded polynomial treats the folds as coefficients and only substitutes values in the monomials of the folded polynomial. Thus for any fixed  $y \in \mathbb{R}^n$ ,  $f(y)$  (sometimes denoted by  $(f(y))(x)$  for contextual clarity) is a degree- $d_2$  polynomial in  $x$ , i.e.  $f(y) \in \mathbb{R}_{d_2}[x]$ .

**Lemma 6.3 (Folded Analogue of Lemma 5.3).**

Let  $f \in (\mathbb{R}_{d_2}^+[x])_{d_1}[x]$  be a  $(d_1, d_2)$ -folded polynomial whose folds have non-negative coefficients. Then for any  $\alpha \in \mathbb{N}_{\leq d_1/2}^n$  and any  $y \geq 0$ ,

$$\left( f \left( y + \frac{\sqrt{\alpha}}{\sqrt{|\alpha|}} \right) \right)(x) \geq \frac{(F_{2\alpha}(y))(x)}{|\mathcal{O}(\alpha)|} \cdot 2^{-O(d_1)}$$

where the ordering is coefficient-wise.

*Proof.* Identical to the proof of Lemma 5.3. ■

**Lemma 6.4 (Folded Analogue of Lemma 5.11).**

Let  $f \in (\mathbb{R}_{d_2}[x])_{d_1}[x]$  be a  $(d_1, d_2)$ -folded polynomial. Consider any  $\alpha \in \mathbb{N}_{\leq d_1/2}^n$  and any  $y$ , and let

$$z := \Xi \cdot y \circ \frac{1}{2\alpha + \mathbb{1}} \circ b + \frac{\sqrt{\alpha} \circ \zeta}{\sqrt{|\alpha|}}$$

where  $\Xi$  is an independent and uniformly randomly chosen  $(d_1 - 2|\alpha| + 1)$ -th root of unity, and for any  $i \in [n]$ ,  $\zeta_i$  is an independent and uniformly randomly chosen  $(2\alpha_i + 1)$ -th root of unity, and  $b_i$  is an independent Bernoulli( $p$ ) random variable ( $p$  is an arbitrary parameter in  $[0, 1]$ ). Then

$$\mathbb{E} \left[ (f(z))(x) \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] = p^{d_1 - 2|\alpha|} \cdot \frac{(F_{2\alpha}(y))(x)}{|\mathcal{O}(\alpha)|} \cdot 2^{-O(d_1)} + r(p)$$

where  $r(p)$  is a univariate polynomial in  $p$  with degree less than  $d_1 - 2|\alpha|$  (and whose coefficients are in  $\mathbb{R}_{d_2}[x]$ ).

*Proof.* This follows by going through the proof of Lemma 5.11 for every fixed  $x$ . ■

## 6.4. Bounding $\Lambda_C()$ of Multilinear Folded Polynomials

Here we bound  $\Lambda_C()$  of a multilinear folded polynomial in terms of properties of the polynomial that are inspired by treating the folds as coefficients and generalizing the coefficient-based approximations for regular (non-folded) polynomials from Theorem 5.15 and Theorem 5.13.

### 6.4.1. General Folds: Bounding $\Lambda()$ in terms of $\Lambda()$ of the "worst" fold

Here we will give a folded analogue of the proof of Theorem 5.15 wherein we used Gershgorin-Circle theorem to bound SOS value in terms of the max-magnitude-coefficient.

**Lemma 6.5 (Folded Analogue of Gershgorin Circle Bound on Spectral Radius).** For even  $d_1, d_2$ , let  $d = d_1 + d_2$ , let  $f \in (\mathbb{R}_{d_2}[x])_{d_1}[x]$  be a multilinear  $(d_1, d_2)$ -folded polynomial. We have,

$$\Lambda(f) \leq 2^{O(d)} \frac{n^{d_1/2}}{d_1^{d_1}} \max_{\gamma \in \{0,1\}_{d_1}^n} \|\bar{f}_\gamma\|_{sp}.$$

*Proof.* Since  $\Lambda(f) \leq \|f\|_{sp}$ , it is sufficient to bound  $\|f\|_{sp}$ .

Let  $M_{\bar{f}_\gamma}$  be the matrix representation of  $\bar{f}_\gamma$  realizing  $\|\bar{f}_\gamma\|_{sp}$ . Let  $M_f$  be an  $[n]^{d_1/2} \times [n]^{d_1/2}$  block matrix with  $[n]^{d_2/2} \times [n]^{d_2/2}$  size blocks, where for any  $I, J \in [n]^{d_1/2}$  the block of  $M_f$  at index  $(I, J)$  is defined to be  $\frac{1}{d_1!} \cdot M_{\bar{f}_{\alpha(I)+\alpha(J)}}$ . Clearly  $M_f$  (interpreted as an  $[n]^{d/2} \times [n]^{d/2}$ ) is a matrix representation of the unfolding of  $f$  since  $f$  is a multilinear folded polynomial. Lastly, applying Block-Gershgorin circle theorem to  $M_f$  and upper bounding the sum of spectral norms over a block row by  $n^{d_1/2}$  times the max term implies the claim. ■

#### 6.4.2. Non-Negative Coefficient Folds: Relating SoS Value to the SoS Value of the $d_1/2$ -collapse

Observe that in the case of a multilinear degree- $d$  polynomial, the  $d/2$ -collapse corresponds (upto scaling) to the sum of a row of the SOS symmetric matrix representation of the polynomial. We will next develop a folded analogue of the proof of Theorem 5.13 wherein we employed Perron-Frobenius theorem to bound SOS value in terms of the  $d/2$ -collapse.

The proof here however, is quite a bit more subtle than in the general case above. This is because one can apply the block-matrix analogue of Gershgorin theorem (due to Feingold et al. [FV<sup>+</sup>62]) to a matrix representation of the folded polynomial (whose spectral norm is an upper bound on  $\Lambda(\cdot)$ ) in the general case. Loosely speaking, this corresponds to bounding  $\Lambda(f)$  in terms of

$$\max_{\gamma \in \{0,1\}_k^n} \sum_{\theta \in \{0,1\}_k^n} \Lambda(\bar{f}_{\gamma+\theta})$$

where  $k = d_1/2$ . This however is not enough in the nnc case as in order to win the  $1/2$  in the exponent, one needs to relate  $\Lambda_C(f)$  to

$$\max_{\gamma \in \{0,1\}_k^n} \Lambda \left( \sum_{\theta \in \{0,1\}_k^n} \bar{f}_{\gamma+\theta} \right).$$

This however, cannot go through Block-Gershgorin since it is **not** true that the spectral norm of a non-negative block matrix is upper bounded by the max over rows of the spectral norm of the sum of blocks in that row. It instead, can only be upper bounded by the max over rows of the sum of spectral norms of the blocks in that row.

To get around this issue, we skip the intermediate step of bounding  $\Lambda_C(f)$  by the spectral norm of a matrix and instead prove the desired relation directly through the use of pseudo-expectation operators. This involved first finding a pseudo-expectation based proof of Gershgorin/Perron-Frobenius bound on spectral radius that generalizes to folded polynomials in the right way.

**Lemma 6.6 (Folded analogue of Perron-Frobenius Bound on Spectral Radius).** *For even  $d_1 = 2k$ , let  $f \in (\mathbb{R}_{d_2}^+[x])_{d_1}[x]$  be a multilinear  $(d_1, d_2)$ -folded polynomial whose folds have non-negative coefficients. Let  $C$  be the system of polynomial constraints given by  $\{\|x\|_2^2 = 1; \forall \beta \in \mathbb{N}_{d_2}^n, x^\beta \geq 0\}$ . We have,*

$$\Lambda_C(f) \leq \max_{\gamma \in \{0,1\}_k^n} \Lambda_C(\bar{g}_\gamma) \cdot \frac{1}{k!}$$

where

$$\bar{g}_\gamma(x) := \overline{C_k(f)}_\gamma = \sum_{\substack{\theta \leq \mathbb{1} - \gamma \\ \theta \in \mathbb{N}_k^n}} \bar{f}_{\gamma+\theta}(x).$$

*Proof.* Consider any pseudo-expectation operator  $\tilde{\mathbf{E}}_C$  of degree at least  $d_1 + d_2$ . Note that since  $\tilde{\mathbf{E}}_C$  satisfies  $\{\forall \beta \in \mathbb{N}_{d_2}^n, x^\beta \geq 0\}$ , by linearity  $\tilde{\mathbf{E}}_C$  must also satisfy  $\{h \geq 0\}$  for any  $h \in \mathbb{R}_{d_2}^+[x]$  - a fact we will use shortly.

Since  $f$  is a multilinear folded polynomial,  $\bar{f}_\alpha$  is only defined when  $0 \leq \alpha \leq \mathbb{1}$ . If  $\alpha \not\leq \mathbb{1}$ , we define  $\bar{f}_\alpha := 0$ . We have,

$$\begin{aligned}
\tilde{\mathbf{E}}_C[f] &= \sum_{\alpha \in \{0,1\}_{d_1}^n} \tilde{\mathbf{E}}_C[\bar{f}_\alpha \cdot x^\alpha] && (f \text{ is a multilinear folded polynomial}) \\
&= \sum_{I \in [n]^k} \sum_{J \in [n]^k} \tilde{\mathbf{E}}_C[\bar{f}_{\alpha(I)+\alpha(J)} \cdot x^I x^J] \cdot \frac{1}{d_1!} && (\text{by multilinearity}) \\
&\leq \sum_{I \in [n]^k} \sum_{J \in [n]^k} \tilde{\mathbf{E}}_C\left[\bar{f}_{\alpha(I)+\alpha(J)} \cdot \frac{(x^I)^2 + (x^J)^2}{2}\right] \cdot \frac{1}{d_1!} && (\tilde{\mathbf{E}}_C \text{ satisfies } \bar{f}_\alpha \geq 0) \\
&= \sum_{I \in [n]^k} \sum_{J \in [n]^k} \tilde{\mathbf{E}}_C[\bar{f}_{\alpha(I)+\alpha(J)} \cdot (x^I)^2] \cdot \frac{1}{d_1!} \\
&= \sum_{I \in [n]^k} \tilde{\mathbf{E}}_C\left[(x^I)^2 \cdot \sum_{J \in [n]^k} \bar{f}_{\alpha(I)+\alpha(J)}\right] \cdot \frac{1}{d_1!} \\
&= \sum_{I \in [n]^k} \tilde{\mathbf{E}}_C\left[(x^I)^2 \cdot \sum_{\substack{\theta \leq \mathbb{1} - \alpha(I) \\ \theta \in \mathbb{N}_k^n}} \bar{f}_{\alpha(I)+\theta}\right] \cdot \frac{k!}{d_1!} && (\text{by multilinearity}) \\
&= \sum_{I \in [n]^k} \tilde{\mathbf{E}}_C[(x^I)^2 \cdot g_{\alpha(I)}] \cdot \frac{k!}{d_1!} \\
&\leq \sum_{I \in [n]^k} \tilde{\mathbf{E}}_C[(x^I)^2] \cdot \Lambda_C(\bar{g}_{\alpha(I)}) \cdot \frac{1}{k!} && (\text{by Lemma 4.2}) \\
&\leq \sum_{I \in [n]^k} \tilde{\mathbf{E}}_C[(x^I)^2] \cdot \max_{\gamma \in \{0,1\}_k^n} \Lambda_C(\bar{g}_\gamma) \cdot \frac{1}{k!} && (\tilde{\mathbf{E}}_C[(x^I)^2] \geq 0) \\
&= \tilde{\mathbf{E}}_C[\|x\|_2^{d_1}] \cdot \max_{\gamma \in \{0,1\}_k^n} \Lambda_C(\bar{g}_\gamma) \cdot \frac{1}{k!} \\
&= \max_{\gamma \in \{0,1\}_k^n} \Lambda_C(\bar{g}_\gamma) \cdot \frac{1}{k!}
\end{aligned}$$

■

We are finally equipped to prove the main results of this section.

## 6.5. $(n/q)^{d/4-1/2}$ -Approximation for Non-negative Coefficient Polynomials

**Theorem 6.7.** Consider any  $f \in \mathbb{R}_d^+[x]$  for  $d \geq 2$ , and any  $q$  divisible by  $2d$ . Let  $C$  be the system of polynomial constraints given by  $\{\|x\|_2^2 = 1; \forall \beta \in \mathbb{N}_{2q/d}^n, x^\beta \geq 0\}$ . Then we have,

$$\frac{\Lambda_C(f^{q/d})^{d/q}}{\|f\|_2} \leq 2^{O(d)} \frac{n^{d/4-1/2}}{q^{d/4-1/2}}.$$

*Proof.* Let  $h$  be any  $(d-2, 2)$ -folded polynomial whose unfolding yields  $f$  and whose folds have non-negative coefficients and let  $s$  be the  $(\bar{q}, 2q/d)$ -folded polynomial given by  $h^{q/d}$  where  $\bar{q} := (d-2)q/d$ . Finally, consider any  $\alpha \in \mathbb{N}_{\leq \bar{q}/2}^n$  and let  $S_{2\alpha}$  be the multilinear component of  $s$  as defined in Lemma 6.1. We will establish that for any  $\gamma \in \{0, 1\}_k^n$  (where  $k := \bar{q}/2 - |\alpha|$ ),

$$\|f\|_2^{q/d} \geq \frac{2^{-O(q)} \cdot \Lambda_C(\overline{C_{\bar{q}/2-|\alpha|}(S_{2\alpha})}_\gamma)}{(\bar{q}/2 - |\alpha|)^{\bar{q}/4-|\alpha|/2} \cdot |\mathcal{O}(\alpha)| \cdot n^{\bar{q}/4-|\alpha|/2}} \quad (6.1)$$

which on combining with the application of Lemma 6.2 to  $s$  and its composition with Lemma 6.6, yields the claim. To elaborate, we apply Lemma 6.2 to  $s$  with  $d_1 = \bar{q}, d_2 = 2q/d$  and then for every  $\alpha \in \mathbb{N}_{\leq \bar{q}/2}^n$  we apply Lemma 6.6 with  $d_1 = \bar{q} - 2|\alpha|, d_2 = 2q/d$ , to get

$$\Lambda_C(f^{q/d}) = \Lambda_C(s) \leq 2^{O(q)} \cdot \max_{\alpha \in \mathbb{N}_{\leq \bar{q}/2}^n} \max_{\gamma \in \{0,1\}_{\bar{q}/2-|\alpha|}^n} \frac{\Lambda_C(\overline{C_{\bar{q}/2-|\alpha|}(S_{2\alpha})}_\gamma)}{(\bar{q}/2 - |\alpha|)! \cdot |\mathcal{O}(\alpha)|}$$

which on combining with Eq. (6.1) yields the claim.

It remains to establish Eq. (6.1). So fix any  $\alpha, \gamma$  satisfying the above conditions. Let  $t := |\alpha|$  and let  $k := \bar{q}/2 - |\alpha|$ . Clearly  $\|f\|_2 \geq f(y/\|y\|_2)$  where  $y := a + z$ , and

$$z := \frac{\mathbf{1}}{\sqrt{n}} + \frac{\gamma}{\sqrt{k}} + \frac{\sqrt{\alpha}}{\sqrt{t}}$$

and  $a$  is the unit vector that maximizes the quadratic polynomial

$$(h(z))(x).$$

Since  $\|y\|_2 = O(1)$ ,  $\|f\|_2 \geq f(y)/2^{O(d)}$ . Now clearly by non-negativity we have,

$$f(y) \geq (h(z))(a) = \|h(z)\|_2$$

Thus we have,

$$\begin{aligned} \|f\|_2^{q/d} &\geq \|(h(z))(x)\|_2^{q/d} \cdot 2^{-O(q)} \\ &= \|h(z)^{q/d}(x)\|_2 \cdot 2^{-O(q)} \\ &= \Lambda_C(h(z)^{q/d}(x)) \cdot 2^{-O(q)} && \text{(SOS exact on powered quadratics)} \\ &= \Lambda_C(s(z)(x)) \cdot 2^{-O(q)} \\ &\geq \Lambda_C(S_{2\alpha}(\mathbf{1}/\sqrt{n} + \gamma/\sqrt{k})(x)) \cdot \frac{2^{-O(q)}}{|\mathcal{O}(\alpha)|} && \text{(by Lemma 4.3 and Lemma 6.3)} \\ &\geq \frac{\Lambda_C(\overline{C_k(S_{2\alpha})}_\gamma)}{k^{k/2} \cdot n^{k/2}} \cdot \frac{2^{-O(q)}}{|\mathcal{O}(\alpha)|} && \text{(by Lemma 4.3, and} \\ & && S_{2\alpha}(\frac{\mathbf{1}}{\sqrt{n}} + \frac{\gamma}{\sqrt{k}}) \geq \overline{C_k(S_{2\alpha})}_\gamma \text{ coefficient-wise)} \end{aligned}$$

which completes the proof since we've established Eq. (6.1). ■



## 6.6. $(n/q)^{d/2-1}$ -Approximation for General Polynomials

**Theorem 6.8.** Consider any  $f \in \mathbb{R}_d^+[x]$  for  $d \geq 2$ , and any  $q$  divisible by  $2d$ . Then we have,

$$\frac{\Lambda(f^{q/d})^{d/q}}{\|f\|_2} \leq 2^{O(d)} \frac{n^{d/2-1}}{q^{d/2-1}}.$$

*Proof.* Let  $h$  be the unique  $(d-2, 2)$ -folded polynomial whose unfolding yields  $f$  and such that for any  $\beta \in \mathbb{N}_{d-2}^n$ , the fold  $\bar{h}_\beta$  of  $h$  is equal upto scaling, to the quadratic form of the corresponding  $(n \times n)$  block of the SOS-symmetric matrix representation  $M_f$  of  $f$ . That is, for any  $I, J \in [n]^{d/2-1}$ , s.t.  $\alpha(I) + \alpha(J) = \beta$ ,

$$\bar{h}_\beta(x) = \frac{x^T M_f[I, J] x}{|\mathcal{O}(\beta)|}.$$

Let  $s$  be the  $(\bar{q}, 2q/d)$ -folded polynomial given by  $h^{q/d}$  where  $\bar{q} := (d-2)q/d$ . Consider any  $\alpha \in \mathbb{N}_{\leq \bar{q}/2}^n$  and  $\gamma \in \{0, 1\}_{\bar{q}-2|\alpha|}^n$ , and let  $S_{2\alpha}$  be the multilinear component of  $s$  as defined in Lemma 6.1. Below the fold (no pun intended), we will show

$$\|f\|_2^{q/d} \geq \frac{2^{-O(q)} \cdot \left\| \overline{(S_{2\alpha})_\gamma} \right\|_{sp}}{(\bar{q} - 2|\alpha|)^{\bar{q}/2-|\alpha|} \cdot |\mathcal{O}(\alpha)|} \quad (6.2)$$

which would complete the proof after applying Lemma 6.2 to  $s$  and composing the result with Lemma 6.5. To elaborate, we apply Lemma 6.2 to  $s$  with  $d_1 = \bar{q}, d_2 = 2q/d$  and then for every  $\alpha \in \mathbb{N}_{\leq \bar{q}/2}^n$  we apply Lemma 6.5 with  $d_1 = \bar{q} - 2|\alpha|, d_2 = 2q/d$ , to get

$$\Lambda(f^{q/d}) = \Lambda(s) \leq 2^{O(q)} \cdot \max_{\alpha \in \mathbb{N}_{\leq \bar{q}/2}^n} \max_{\gamma \in \{0, 1\}_{\bar{q}-2|\alpha|}^n} \frac{\left\| \overline{(S_{2\alpha})_\gamma} \right\|_{sp}}{(\bar{q} - 2|\alpha|)^{\bar{q}/2-|\alpha|} \cdot |\mathcal{O}(\alpha)|}$$

which on combining with Eq. (6.2) yields the claim.

Fix any  $\alpha, \gamma$  satisfying the above conditions. Let  $k := \bar{q} - 2\alpha$ . Let  $t := |\alpha|$ , and let

$$z := \Xi \cdot \frac{1}{\sqrt{k}} \cdot \gamma \circ \frac{1}{2\alpha + \mathbb{1}} \circ b + \frac{\sqrt{\alpha} \circ \zeta}{\sqrt{t}}$$

$\Xi$  is an independent and uniformly randomly chosen  $(k+1)$ -th root of unity, and for any  $i \in [n]$ ,  $\zeta_i$  is an independent and uniformly randomly chosen  $(2\alpha_i + 1)$ -th root of unity, and for any  $i \in [n]$ ,  $b_i$  is an independent Bernoulli( $p$ ) random variable ( $p$  is a parameter that will be set later). By Lemma 5.7 and definition of  $h$ , we see that for any  $y$ ,  $\|f\|_2^c \geq \|(h(y))(x)\|_2^c$ . Thus we have,

$$\begin{aligned} \|f\|_2^{q/d} &= \|f^{q/d}\|_2 \\ &\geq \|f^{q/d}\|_2^c \cdot 2^{-O(q)} && \text{(by Lemma 5.8)} \\ &\geq \max_{p \in [0, 1]} \mathbb{E} \left[ \|h(z)^{q/d}(x)\|_2 \right] \cdot 2^{-O(q)} && \text{(by Lemma 5.7)} \\ &= \max_{p \in [0, 1]} \mathbb{E} \left[ \|h(z)^{q/d}(x)\|_{sp} \right] \cdot 2^{-O(q)} && \text{(SOS exact on powered quadratics)} \\ &= \max_{p \in [0, 1]} \mathbb{E} \left[ \|h(z)^{q/d}(x) \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i\|_{sp} \right] \cdot 2^{-O(q)} \end{aligned}$$

$$\begin{aligned}
&\geq \max_{p \in [0,1]} \left\| \mathbb{E} \left[ h(z)^{q/d}(x) \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] \right\|_{sp} \cdot 2^{-O(q)} \\
&= \max_{p \in [0,1]} \left\| \mathbb{E} \left[ (s(z))(x) \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] \right\|_{sp} \cdot 2^{-O(q)} \\
&= \max_{p \in [0,1]} \left\| p^k \cdot \frac{(S_{2\alpha}(\gamma/\sqrt{k}))(x)}{|\mathcal{O}(\alpha)|} + r(p) \right\|_{sp} \cdot 2^{-O(q)} \quad (\text{by Lemma 6.4, } \deg(r) < k) \\
&= \max_{p \in [0,1]} \left\| p^k \cdot \frac{(\overline{S_{2\alpha}})_\gamma(x)}{k^{k/2} \cdot |\mathcal{O}(\alpha)|} + r(p) \right\|_{sp} \cdot 2^{-O(q)} \\
&\geq \frac{\|(\overline{S_{2\alpha}})_\gamma\|_{sp}}{k^{k/2} \cdot |\mathcal{O}(\alpha)|} \cdot 2^{-O(q+k)} \quad (\text{Chebyshev Inequality - Lemma 5.5})
\end{aligned}$$

where the last inequality follows by the following argument: one would like to show that there always exists  $p \in [0, 1]$  such that  $\|p^k \cdot h_k(x) + \dots p^0 \cdot h_0(x)\|_{sp} \geq \|h_k(x)\|_{sp} \cdot 2^{-O(k)}$ . So let  $p$  be such that  $|p^k \cdot u^T M_k v + \dots p^0 \cdot u^T M_0 v| \geq |u^T M_k v| \cdot 2^{-O(k)}$  (such a  $p$  exists by Chebyshev inequality) where  $M_k$  is the matrix representation of  $h_k(x)$  realizing  $\|h_k\|_{sp}$  and  $u, v$  are the maximum singular vectors of  $M_k$ .  $M_{k-1}, \dots, M_0$  are arbitrary matrix representations of  $h_{k-1}, \dots, h_0$  respectively. But  $p^k \cdot M_k + \dots p^0 \cdot M_0$  is a matrix representation of  $p^k \cdot h_k + \dots p^0 \cdot h_0$ . Thus  $\|p^k \cdot h_k + \dots p^0 \cdot h_0\|_{sp} \geq |u^T M_k v| \cdot 2^{-O(k)} = \|h_k\|_{sp} \cdot 2^{-O(k)}$ .

This completes the proof as we've established Eq. (6.2). ■

## 6.7. Algorithms

It is straightforward to extract algorithms from the proofs of Theorem 6.7 and Theorem 6.8.

### 6.7.1. Non-negative coefficient polynomials

Let  $f$  be a degree- $d$  polynomial with non-negative coefficients and let  $h$  be a  $(d-2, 2)$ -folded polynomial whose unfolding yields  $f$ . Consider any  $q$  divisible by  $2d$  and let  $\bar{q} := (d-2)q/d$ . Pick and return the best vector from the set

$$\left\{ \frac{\mathbb{1}}{\sqrt{n}} + \frac{\sqrt{\alpha}}{\sqrt{|\alpha|}} + \frac{\gamma}{\sqrt{|\gamma|}} + \arg \max \left\| h \left( \frac{\mathbb{1}}{\sqrt{n}} + \frac{\sqrt{\alpha}}{\sqrt{|\alpha|}} + \frac{\gamma}{\sqrt{|\gamma|}} \right) (x) \right\|_2 \mid \alpha \in \mathbb{N}_{\leq \bar{q}/2}^n, \gamma \in \mathbb{N}_{\bar{q}/2-|\alpha|}^n \right\}$$

### 6.7.2. General Polynomials

Let  $f$  be a degree- $d$  polynomial and let  $h$  be the unique  $(d-2, 2)$ -folded polynomial whose unfolding yields  $f$  and such that for any  $\beta \in \mathbb{N}_{d-2}^n$ , the fold  $\bar{h}_\beta$  of  $h$  is equal upto scaling, to the quadratic form of the corresponding  $(n \times n)$  block of the SOS-symmetric matrix representation  $M_f$  of  $f$ . That is, for any  $I, J \in [n]^{d/2-1}$ , s.t.  $\alpha(I) + \alpha(J) = \beta$ ,

$$\bar{h}_\beta(x) = \frac{x^T M_f[I, J] x}{|\mathcal{O}(\beta)|}.$$

Consider any  $q$  divisible by  $2d$  and let  $\bar{q} := (d-2)q/d$ . Let the set  $S$  be defined by,

$$S := \left\{ \Xi \cdot \frac{1}{\sqrt{|\gamma|}} \cdot \gamma \circ \frac{1}{2\alpha + \mathbb{1}} \circ b + \frac{\sqrt{\alpha} \circ \zeta}{\sqrt{|\alpha|}} \mid \begin{array}{l} \Xi \in \Omega_{k+1}, \zeta_i \in \Omega_{2\alpha_i+1}, b \in \{0,1\}^n, \\ \alpha \in \mathbb{N}_{\leq \bar{q}/2}^n, \gamma \in \{0,1\}_{\bar{q}-2|\alpha|}^n \end{array} \right\}$$

where  $\Omega_p$  denotes the set of  $p$ -th roots of unity. Pick and return the best vector from the set

$$\left\{ c_1 \cdot y + c_2 \cdot \arg \max \| (h(y))(x) \|_2 \mid y \in S, c_1 \in [-(d-2), (d-2)], c_2 \in [-2, 2] \right\}$$

Note that one need only search through all roots of unity vectors  $\zeta$  supported on  $S(\gamma)$  and all  $\{0,1\}$ -vectors  $b$  supported on  $S(\alpha)$ . Lemma 5.7 can trivially be made constructive in time  $2^{O(q)}$ . Lastly, to go from complexes to reals, Lemma 5.8 can trivially be made constructive using  $2^{O(d)}$  time. Thus the algorithm runs in time  $n^{O(q)}$ .

## 7. Random Polynomials

### 7.1. Upper Bounds

Let  $A \in \mathbb{R}^{[n]^{d/2} \times [n]^{d/2}}$  be a random matrix whose entry is an independent Rademacher random variable, and let  $f := \sum_{I,J \in [n]^{d/2}} A[I,J] \cdot x^{\alpha(I) + \alpha(J)}$ . With high probability  $\|f\|_2 = O(\sqrt{n} \log n)$ . In this section, we prove that for every  $q$  divisible by  $d$ , with high probability,

$$\left( \Lambda(f^{q/d}) \right)^{d/q} \leq \tilde{O}_d \left( \left( \frac{n}{q^{1-2/d}} \right)^{d/4} \right) = \tilde{O}_d \left( \left( \frac{n}{q} \right)^{d/4-1/2} \right) \cdot \|f\|_2.$$

To prove it, we use the following matrix representation  $M$  of  $f^{q/d}$ , and show that  $\|M\|_2 \leq \tilde{O}_d \left( \left( \frac{n \log^5 n}{q^{1-2/d}} \right)^{q/4} \right)$ . Given a tuple  $I = (i_1, \dots, i_q)$ , and an integer  $d$  that divides  $q$  and  $1 \leq \ell \leq q/d$ , let  $I_{\ell;d}$  be the  $d$ -tuple  $(I_{d(\ell-1)+1}, \dots, I_{d\ell})$  (i.e., if we divide  $I$  into  $q/d$  tuples of length  $d$ ,  $I_{\ell;d}$  be the  $\ell$ th tuple). Furthermore, given a tuple  $I = (i_1, \dots, i_q) \in [n]^q$  and a permutation  $\pi \in [n]^q$ , let  $\pi(I)$  be another  $q$ -tuple whose  $\ell$ th coordinate is  $\pi(i_\ell)$ . For  $I, J \in [n]^{q/2}$ ,  $M[I, J]$  is formally given by

$$\begin{aligned} M[I, J] &= \frac{1}{q!} \cdot \sum_{\pi, \sigma \in \mathbb{S}_{q/2}} A^{\otimes q/d}[\pi(I), \sigma(J)] \\ &= \frac{1}{q!} \cdot \sum_{\pi, \sigma \in \mathbb{S}_{q/2}} \prod_{\ell=1}^{q/d} A[(\pi(I))_{\ell;d/2}, (\sigma(J))_{\ell;d/2}]. \end{aligned}$$

We perform the trace method to bound  $\|M\|_2$ . Let  $p$  be an even integer, that will be eventually taken as  $\Theta(\log n)$ .  $\text{Tr}(M)$  can be written as (let  $I^{p+1} := I^1$ )

$$\mathbb{E} \left[ \sum_{I^1, \dots, I^p \in [n]^{q/2}} \prod_{\ell=1}^p M[I^\ell, I^{\ell+1}] \right] = \sum_{I^1, \dots, I^p} \mathbb{E} \left[ \prod_{\ell=1}^p \left( \sum_{\pi_j, \sigma_j \in \mathbb{S}_{q/2}} \prod_{m=1}^{q/d} A[(\pi(I^\ell))_{m;d/2}, (\sigma(I^{\ell+1}))_{m;d/2}] \right) \right].$$

Let  $E(I^1, \dots, I^p) := \mathbb{E} [\prod_{\ell=1}^p M[I^\ell, I^{\ell+1}]]$ , which is the expected value in the right hand side. To analyze  $E(I^1, \dots, I^p)$ , we first introduce notions to classify  $I^1, \dots, I^p$  depending on their intersection patterns. For any  $I^1, \dots, I^p \in [n]^{q/2}$ , let  $e_k$  denote the  $k$ -th smallest element in  $\bigcup_{\ell,j} \{i_j^\ell\}$ . For any  $c^1, \dots, c^s \in [q/2]^p$ , let

$$\mathcal{C}(c^1 \dots c^s) := \left\{ (I^1, \dots, I^p) \mid \#(I^1, \dots, I^p) = s, \forall k \in [s], \ell \in [p], e_k \text{ appears } c_\ell^k \text{ times in } I^\ell \right\}.$$

The following two observations on  $c^1, \dots, c^s$  can be easily proved.

**Observation 7.1.** *If  $\mathcal{C}(c^1, \dots, c^s) \neq \emptyset$ ,*

$$|\mathcal{C}(c^1, \dots, c^s)| \leq \frac{n^s}{s!} \times \frac{((q/2)!)^p}{\prod_{\ell \in [p]} c_\ell^1! \dots c_\ell^s!}.$$

Moreover,

$$\left| \left\{ (c^1, \dots, c^s) \in ([q/2]^p)^s \mid \mathcal{C}(c^1, \dots, c^s) \neq \emptyset \right\} \right| \leq 2^{O(pq)} p^{pq/2}.$$

The following lemma bounds  $E(I^1, \dots, I^p)$  in terms of the corresponding  $c_1, \dots, c_s$ .

**Lemma 7.2.** *Consider any  $c^1, \dots, c^s \in [q/2]^p$  and  $(I^1, \dots, I^p) \in \mathcal{C}(c^1, \dots, c^s)$ . We have*

$$E(I^1, \dots, I^p) \leq 2^{O(pq)} \frac{p^{1/2+1/2d}}{q^{1/2-1/2d}} \prod_{\ell \in [p]} c_\ell^1! \dots c_\ell^s!$$

*Proof.* Consider any  $c^1, \dots, c^s \in [q/2]^p$  and  $(I^1, \dots, I^p) \in \mathcal{C}(c^1, \dots, c^s)$ . We have

$$\begin{aligned} & E(I^1, \dots, I^p) \\ &= \mathbb{E} \left[ \prod_{\ell=1}^p M[I^\ell, I^{\ell+1}] \right] \\ &= \sum_{\pi_j, \sigma_j \in \mathbb{S}_{q/2}} \mathbb{E} \left[ \prod_{\ell=1}^p \prod_{m=1}^{q/d} A[(\pi(I^\ell))_{m;d/2}, (\pi(I^{\ell+1}))_{m;d/2}] \right] \\ &= \left( \frac{\prod_{\ell} \prod_s (c_\ell^s!)^2}{((q/2)!)^{2p}} \right) \cdot \sum_{(J^\ell, K^\ell \in \mathcal{O}(I^\ell))_{\ell \in [p]}} \mathbb{E} \left[ \prod_{\ell=1}^p \prod_{m=1}^{q/d} A[J_{m;d/2}^\ell, K_{m;d/2}^{\ell+1}] \right] \end{aligned} \quad (7.1)$$

Thus,  $E(I^1, \dots, I^p)$  is bounded by the number of choices for  $J^1, \dots, J^p, K^1, \dots, K^p$  such that  $J^\ell, K^\ell \in \mathcal{O}(I^\ell)$  for each  $\ell \in [p]$ , and  $\mathbb{E} \left[ \prod_{\ell=1}^p \prod_{m=1}^{q/d} A[J_{m;d/2}^\ell, K_{m;d/2}^{\ell+1}] \right]$  is nonzero.

Given  $J^1, \dots, J^p$  and  $K^1, \dots, K^p$ , consider the  $(pq/d)$ -tuple  $T$  where each coordinate is indexed by  $(\ell, m)_{\ell \in [p], m \in [q/d]}$  and has a  $d$ -tuple  $T_{\ell, m} := (J_{m;d/2}^\ell) \oplus (K_{m;d/2}^{\ell+1}) \in \mathbb{R}^d$  as a value. Note that  $\sum_{\ell, m} \alpha(T_{\ell, m}) = (2o_1, \dots, 2o_n)$  where  $o_r$  is the number of occurrences of  $r \in [n]$  in  $(pq/2)$ -tuple  $\oplus_{\ell=1}^p I^\ell$ . The fact that  $\mathbb{E} \left[ \prod_{\ell=1}^p \prod_{m=1}^{q/d} A[j_{m;d/2}, k_{m;d/2}] \right] \neq 0$  means that every  $d$ -tuple occurs even number of times in  $T$ .

We count the number of  $(pq/d)$ -tuples  $T = (T_{\ell, m})_{\ell \in [p], m \in [q/d]}$  that  $\sum_{\ell, m} \alpha(T_{\ell, m}) = (2o_1, \dots, 2o_n)$  and every  $d$ -tuple occurs exactly even number of times. Let  $Q = (Q_1, \dots, Q_{pq/2d}), R = (R_1, \dots, R_{pq/2d})$

be two  $(pq/2d)$ -tuples of  $d$ -tuples where for every  $d$ -tuple  $P$ , the number of occurrences of  $P$  is the same in  $Q$  and  $R$ , and  $\sum_{\ell=1}^{pq/2d} \alpha(Q_\ell) = \sum_{\ell=1}^{pq/2d} \alpha(R_\ell) = (o_1, \dots, o_n)$ . At most  $2^{pq/d}$  tuples  $T$  can be made by interleaving  $Q$  and  $R$  — for each  $(\ell, m)$ , choose  $T_{\ell,m}$  from the first unused  $d$ -tuple in either  $Q$  or  $R$ . Furthermore, every tuple  $T$  that meets our condition can be constructed in this way.

Due to the condition  $\sum_{\ell=1}^{pq/2d} \alpha(Q_\ell) = (o_1, \dots, o_n)$ , the number of choices for  $Q$  is at most the number of different ways to permute  $I^1 \oplus \dots \oplus I^p$ , which is at most  $(pq/2)! / \prod_{m \in [s]} (\bar{c}^m)!$ , where  $\bar{c}^m := \sum_{\ell \in [p]} c_\ell^m$  for  $m \in [s]$ . For a fixed choice of  $Q$ , there are at most  $(pq/2d)!$  choices of  $R$ . Therefore, the number of choices for  $(J^\ell, K^\ell \in \mathcal{O}(I^\ell))_{\ell \in [p]}$  with nonzero expected value is at most

$$2^{pq/d} \cdot \frac{(pq/2)!}{\prod_{m \in [s]} (\bar{c}^m)!} \cdot (pq/2d)! = 2^{O(pq)} \cdot \frac{(pq)^{1/2+1/2d}}{\prod_{m \in [s]} (\bar{c}^m)!}.$$

Combining with Eq. (7.1),

$$E(I^1, \dots, I^p) \leq \left( 2^{O(pq)} \frac{(pq)^{1/2+1/2d}}{\prod_{m \in [s]} (\bar{c}^m)!} \right) \cdot \left( \frac{\prod_{\ell} \prod_s (c_\ell^s!)^2}{((q/2)!)^{2p}} \right) \leq 2^{O(pq)} \cdot \frac{p^{1/2+1/2d}}{q^{1/2-1/2d}} \cdot \prod_{\ell} \prod_s c_\ell^s!$$

as desired.  $\blacksquare$

**Lemma 7.3.** For all  $I^1, \dots, I^p \in [n]^{q/2}$ , if  $E(I^1, \dots, I^p) \neq 0$ ,  $\#(I^1, \dots, I^p) \leq \frac{pq}{4} + \frac{q}{2}$ .

*Proof.* Note that  $E(I^1, \dots, I^p) \neq 0$  implies that there exist  $J^1, \dots, J^p, K^1, \dots, K^p$  such that  $J^\ell, K^\ell \in \mathcal{O}(I^\ell)$  and every  $d$ -tuple occurs exactly even number of times in  $((J_{m;d/2}^\ell) \oplus (K_{m;d/2}^{\ell+1}))_{\ell \in [p], m \in [q/2]}$ . Consider the graph  $G = (V, E)$  defined by

$$V := \bigcup_{\ell \in [p]} \bigcup_{k \in [q/2]} \{I_k^\ell\}$$

$$E := \bigcup_{m \in [q/2]} \left\{ \{J_m^1, K_m^2\}, \{J_m^2, K_m^3\}, \dots, \{J_m^p, K_m^1\} \right\}.$$

The even multiplicity condition implies that every element in  $E$  has even multiplicity and consequently  $|E| \leq pq/4$ . We next show that  $E$  is the union of  $q/2$  paths. To this end, we construct  $G^1 \in \mathcal{O}(I^1), \dots, G^\ell \in \mathcal{O}(I^\ell)$  as follows:

1. Let  $G^2 := K^2$
2. For  $3 \leq \ell \leq p$  do:

Since  $G^\ell \in \mathcal{O}(I^\ell)$ , there exists  $\pi \in \mathbb{S}_{q/2}$  s.t.  $\pi(J^\ell) = G^\ell$ .

Let  $G^{\ell+1} := \pi(K^{\ell+1})$ .

We observe that by construction,

$$\bigcup_{m \in [q/2]} \left\{ \{J_m^1, G_m^2\}, \{G_m^2, G_m^3\}, \dots, \{G_m^p, G_m^1\} \right\} = \bigcup_{m \in [q/2]} \left\{ \{J_m^1, K_m^2\}, \{J_m^2, K_m^3\}, \dots, \{J_m^p, K_m^1\} \right\} = E$$

which establishes that  $E$  is a union of  $q/2$  paths.

Now since  $E$  is the union of  $q/2$  paths  $G$  has at most  $q/2$  connected components, and one needs to add at most  $q/2 - 1$  edges make it connected, we have  $|V| \leq |E| + (q/2 - 1) + 1 \leq pq/4 + q/2$ . But  $\#(I^1, \dots, I^p) = |V|$ , which completes the proof.  $\blacksquare$

Finally,  $\mathbb{E} [\text{Tr}(M^p)]$  can be bounded as follows.

$$\begin{aligned}
& \mathbb{E} [\text{Tr}(M^p)] \\
&= \sum_{I^1, \dots, I^p \in [n]^{q/2}} E(I^1, \dots, I^p) \\
&= \sum_{s \in [pq/4+q/2]} \sum_{\#(I^1, \dots, I^p)=s} E(I^1, \dots, I^p) \quad (\text{by Lemma 7.3}) \\
&= \sum_{s \in [pq/4+q/2]} \sum_{c^1, \dots, c^s \in [q/2]^p} \sum_{(I^1, \dots, I^p) \in \mathcal{C}(c^1 \dots c^s)} E(I^1, \dots, I^p) \\
&= \sum_{s \in [pq/4+q/2]} \sum_{c^1, \dots, c^s \in [q/2]^p} \sum_{(I^1, \dots, I^p) \in \mathcal{C}(c^1 \dots c^s)} E(I^1, \dots, I^p) \\
&\leq \sum_{s \in [pq/4+q/2]} \sum_{c^1, \dots, c^s \in [q/2]^p} \sum_{(I^1, \dots, I^p) \in \mathcal{C}(c^1 \dots c^s)} 2^{O(pq)} \frac{p^{(1/2+1/2d)pq}}{q^{(1/2-1/2d)pq}} \prod_{\ell \in [p]} c_\ell^1! \dots c_\ell^s! \quad (\text{by Lemma 7.2}) \\
&\leq \sum_{s \in [pq/4+q/2]} 2^{O(pq)} \frac{n^s}{s!} p^{(1+1/2d)pq} q^{pq/2d} \quad (\text{by Observation 7.1}) \\
&\leq \sum_{s \in [pq/4+q/2]} 2^{O(pq)} \frac{n^{pq/4+q/2}}{s! q^{pq/4+q/2-s}} p^{(1/2+1/2d)p1} q^{(1/2-1/2d)pq} \quad (\text{assuming } q \leq n) \\
&\leq \sum_{s \in [pq/4+q/2]} 2^{O(pq)} \frac{n^{pq/4+q/2} p^{(1+1/2d)pq}}{q^{(1/4-1/2d)pq}} \\
&\leq 2^{O(pq)} \frac{n^{pq/4+q/2} p^{(1+1/2d)pq}}{q^{(1/4-1/2d)pq}}.
\end{aligned}$$

Choose  $p$  to be even and let  $p = \Theta(\log n)$ . Applying Markov inequality shows that with high probability,

$$\left( \Lambda(f^{q/d}) \right)^{d/q} \leq (\|M\|_2)^{d/q} \leq (\mathbb{E} [\text{Tr}(M^p)])^{d/pq} = O\left(\frac{n^{d/4} \log^{(1+1/2d)} n}{q^{d/4-1/2}}\right) = \tilde{O}\left(\frac{n^{d/4}}{q^{d/4-1/2}}\right).$$

## 7.2. Lower Bounds

Let  $A \in \mathbb{R}^{[n]^{d/2} \times [n]^{d/2}}$  be a random matrix whose entry is an independent Rademacher random variable, and let  $f := \sum_{i,j \in [n]^{d/2}} A[i, j] \cdot x^{\alpha(i) + \alpha(j)}$ . This section proves Theorem 1.3, by constructing a moment matrix  $M$  that is positive semidefinite, SoS-symmetric,  $\text{Tr}(M) = 1$ , and  $\langle A, M \rangle \geq 2^{-O(d)} \cdot \frac{n^{d/4}}{d^{d/4}}$ . In Section 7.2.1, we construct the matrix  $\hat{W}$  that acts as a SoS-symmetrized identity matrix. The moment matrix  $M$  is presented in Section 7.2.2.

### 7.2.1. Wigner Moment Matrix

In this section, we construct an SoS-symmetric and positive semidefinite matrix  $\hat{W} \in \mathbb{R}^{\mathbb{N}_{q/2}^n \times \mathbb{N}_{q/2}^n}$  such that  $\lambda_{\min}(\hat{W}) / \text{Tr}(\hat{W}) \geq 1/(2^{q+1} \cdot |\mathbb{N}_{q/2}^n|)$ , i.e. the ratio of the minimum eigenvalue to the average eigenvalue is at least  $1/2^{q+1}$ .

**Theorem 7.4.** *For any positive integer  $n$  and any positive even integer  $q$ , there exists a matrix  $\hat{W} \subseteq \mathbb{R}^{\mathbb{N}_{q/2}^n \times \mathbb{N}_{q/2}^n}$  that satisfies the following three properties: (1)  $\hat{W}$  is degree- $q$  SoS symmetric. (2) The minimum eigenvalue of  $\hat{W}$  is at least  $\frac{1}{2}$ . (3) Each entry of  $\hat{W}$  is in  $[0, 2^q]$ .*

Theorem 7.4 is proved by explicitly constructing independent random variables  $x_1, \dots, x_n$  such that for any  $n$ -variate polynomial  $p(x_1, \dots, x_n)$  of degree at most  $\frac{q}{2}$ ,  $\mathbb{E}[p^2]$  is bounded away from 0. The proof consists of three parts. The first part shows the existence of a desired distribution for one variable  $x_i$ . The second part uses induction to prove that  $\mathbb{E}[p^2]$  is bounded away from 0. The third part constructs  $\widehat{W} \subseteq \mathbb{R}^{\mathbb{N}_{q/2}^n \times \mathbb{N}_{q/2}^n}$  from the distribution defined.

**Wigner Semicircle Distribution and Hankel Matrix.** Let  $k$  be a positive integer. In this part, the rows and columns of all  $(k+1) \times (k+1)$  matrices are indexed by  $\{0, 1, \dots, k\}$ . Let  $T$  be a  $(k+1) \times (k+1)$  matrix where  $T[i, j] = 1$  if  $|i - j| = 1$  and  $T[i, j] = 0$  otherwise. Let  $e_0 \in \mathbb{R}^{k+1}$  be such that  $(e_0)_0 = 1$  and  $(e_0)_i = 0$  for  $1 \leq i \leq k$ . Let  $R \in \mathbb{R}^{(k+1) \times (k+1)}$  be defined by  $R := [e_0, Te_0, T^2e_0, \dots, T^ke_0]$ . Let  $R_0, \dots, R_k$  be the columns of  $R$  so that  $R_i = T^ie_0$ . It turns out that  $R$  is closely related to the number of ways to consistently put parentheses. Given a string of parentheses ‘(’ or ‘)’, we call it *consistent* if any prefix has at least as many ‘(’ as ‘)’. For example,  $((()))$  is consistent, but  $((()))$  is not.

**Claim 7.5.**  $R[i, j]$  is the number of ways to place  $j$  parentheses ‘(’ or ‘)’ consistently so that there are  $i$  more ‘(’ than ‘)’.

*Proof.* We proceed by the induction on  $j$ . When  $j = 0$ ,  $R[0, 0] = 1$  and  $R[i, 0] = 0$  for all  $i \geq 1$ . Assume the claim holds up to  $j - 1$ . By the definition  $R_j = TR_{j-1}$ .

- For  $i = 0$ , the last parenthesis must be the close parenthesis, so the definition  $R[0, j] = R[1, j - 1]$  still measures the number of ways to place  $j$  parentheses with equal number of ‘(’ and ‘)’.
- For  $i = k$ , the last parenthesis must be the open parenthesis, so the definition  $R[k, j] = R[k - 1, j - 1]$  still measures the number of ways to place  $j$  parentheses with  $k$  more ‘(’.
- For  $0 < i < k$ , the definition of  $R$  gives  $R[i, j] = R[i - 1, j - 1] + R[i + 1, j - 1]$ . Since  $R[i - 1, j]$  corresponds to placing ‘)’ in the  $j$ th position and  $R[i + 1, j]$  corresponds to placing ‘(’ in the  $j$ th position,  $R[i, j]$  still measures the desired quantity.

This completes the induction and proves the claim. ■

Easy consequences of the above claim are (1)  $R[i, i] = 1$  for all  $0 \leq i \leq k$ , and  $R[i, j] = 0$  for  $i > j$ , and (2)  $R[i, j] = 0$  if  $i + j$  is odd, and  $R[i, j] \geq 1$  if  $i \leq j$  and  $i + j$  is even.

Let  $H := (R^T)R$ . Since  $R$  is upper triangular with 1’s on the main diagonal,  $H = (R^T)R$  gives the unique Cholesky decomposition, so  $H$  is positive definite. It is easy to see that  $H[i, j] = \langle R_i, R_j \rangle$  is the total number of ways to place  $i + j$  parentheses consistently with the same number of ‘(’ and ‘)’. Therefore,  $H[i, j] = 0$  if  $i + j$  is odd, and if  $i + j$  is even (let  $l := \frac{i+j}{2}$ ),  $H[i, j]$  is the  $l$ th Catalan number  $C_l := \frac{1}{l+1} \binom{2l}{l}$ . In particular,  $H[i, j] = H[i', j']$  for all  $i + j = i' + j'$ . Such  $H$  is called a *Hankel matrix*.

Given a sequence of  $m_0 = 1, m_1, m_2, \dots$  of real numbers, the *Hamburger moment problem* asks whether there exists a random variable  $W$  supported on  $\mathbb{R}$  such that  $\mathbb{E}[W^i] = m_i$ . It is well-known that there exists a unique such  $W$  if for all  $k \in \mathbb{N}$ , the Hankel matrix  $H \in \mathbb{R}^{(k+1) \times (k+1)}$  defined by  $H[i, j] := \mathbb{E}[W^{i+j}]$  is positive definite [Sim98]. Since our construction of  $H \in \mathbb{R}^{(k+1) \times (k+1)}$  ensures its positive definiteness for any  $k \in \mathbb{N}$ , there exists a unique random variable  $W$  such that  $\mathbb{E}[W^i] = 0$  if  $i$  is odd,  $\mathbb{E}[W^i] = C_{\frac{i}{2}}$  if  $i$  is even. It is known as the *Wigner semicircle distribution* with radius  $R = 2$ .



**Remark 7.6.** Some other distributions (e.g., Gaussian) will give an asymptotically weaker bound. Let  $G$  be a standard Gaussian random variable. The quantitative difference comes from the fact that  $\mathbb{E}[W^{2l}] = C_l = \frac{1}{l+1} \binom{2l}{l} \leq 2^l$  while  $\mathbb{E}[G^{2l}] = (2l-1)!! \geq 2^{\Omega(l \log l)}$ .

**Multivariate Distribution.** Fix  $n$  and  $q$ . Let  $k = \frac{q}{2}$ . Let  $H \in \mathbb{R}^{(k+1) \times (k+1)}$  be the Hankel matrix defined as above, and  $W$  be a random variable sampled from the Wigner semicircle distribution. Consider  $x_1, \dots, x_n$  where each  $x_i$  is an independent copy of  $\frac{W}{N}$  for some large number  $N$  to be determined later. Our  $\widehat{W}$  is later defined to be  $\widehat{W}[\alpha, \beta] = \mathbb{E}[x^{\alpha+\beta}] \cdot N^q$  so that the effect of the normalization by  $N$  is eventually cancelled, but large  $N$  is needed to prove the induction that involves non-homogeneous polynomials.

We study  $\mathbb{E}[p(x)^2]$  for any  $n$ -variate (possibly non-homogeneous) polynomial  $p$  of degree at most  $k$ . For a multivariate polynomial  $p = \sum_{\alpha \in \mathbb{N}_{\leq k}^n} p_\alpha x^\alpha$ , define  $\ell_2$  norm of  $p$  to be  $\|p\|_{\ell_2} := \sqrt{\sum_\alpha p_\alpha^2}$ . For  $0 \leq m \leq n$  and  $0 \leq l \leq k$ , let  $\sigma(m, l) := \inf_p \mathbb{E}[p(x)^2]$  where the infimum is taken over polynomials  $p$  such that  $\|p\|_{\ell_2} = 1$ ,  $\deg(p) \leq l$ , and  $p$  depends only on  $x_1, \dots, x_m$ .

**Lemma 7.7.** There exists  $N := N(n, k)$  such that  $\sigma(m, l) \geq \frac{(1-\frac{m}{2n})}{N^{2l}}$  for all  $0 \leq m \leq n$  and  $0 \leq l \leq k$ .

*Proof.* We prove the lemma by induction on  $m$  and  $l$ . When  $m = 0$  or  $l = 0$ ,  $p$  becomes the constant polynomial 1 or  $-1$ , so  $\mathbb{E}[p^2] = 1$ .

Fix  $m, l > 0$  and a polynomial  $p = p(x_1, \dots, x_m)$  of degree at most  $l$ . Decompose  $p = \sum_{i=0}^l p_i x_m^i$  where each  $p_i$  does not depend on  $x_m$ . The degree of  $p_i$  is at most  $l - i$ .

$$\mathbb{E}[p^2] = \mathbb{E}\left[\left(\sum_{i=0}^l p_i x_m^i\right)^2\right] = \sum_{0 \leq i, j \leq l} \mathbb{E}[p_i p_j] \mathbb{E}[x_m^{i+j}].$$

Let  $\Sigma = \text{diag}(1, \frac{1}{N}, \dots, \frac{1}{N^l}) \in \mathbb{R}^{(l+1) \times (l+1)}$ . Let  $H_l \in \mathbb{R}^{(l+1) \times (l+1)}$  be the submatrix of  $H$  with the first  $l+1$  rows and columns. The rows and columns of  $(l+1) \times (l+1)$  matrices are still indexed by  $\{0, \dots, l\}$ . Define  $R_l \in \mathbb{R}^{(l+1) \times (l+1)}$  similarly from  $R$ , and  $r_t$  ( $0 \leq t \leq l$ ) be the  $t$ th column of  $(R_l)^T$ . Note  $H_l = (R_l)^T R_l = \sum_{t=0}^l r_t r_t^T$ . Let  $H' = \Sigma H_l \Sigma$  such that  $H'[i, j] = \mathbb{E}[x_m^{i+j}]$ . Finally, let  $P \in \mathbb{R}^{(l+1) \times (l+1)}$  be defined such that  $P[i, j] := \mathbb{E}[p_i p_j]$ . Then  $\mathbb{E}[p^2]$  is equal to

$$\text{Tr}(PH') = \text{Tr}(P\Sigma H_l \Sigma) = \text{Tr}\left(P\Sigma\left(\sum_{t=0}^l r_t r_t^T\right)\Sigma\right) = \sum_{t=0}^l \mathbb{E}\left[\left(p_t \frac{1}{N^t} + p_{t+1} \frac{(r_t)_{t+1}}{N^{t+1}} + \dots + p_l \frac{(r_t)_l}{N^l}\right)^2\right],$$

where the last step follows from the fact that  $(r_t)_j = 0$  if  $j < t$  and  $(r_t)_t = 1$ . Consider the polynomial

$$q_t := p_t \frac{1}{N^t} + p_{t+1} \frac{(r_t)_{t+1}}{N^{t+1}} + \dots + p_l \frac{(r_t)_l}{N^l}.$$

Since  $p_i$  is of degree at most  $l - i$ ,  $q_t$  is of degree at most  $l - t$ . Also recall that each entry of  $R$  is bounded by  $2^k$ . By the triangle inequality,

$$\|q_t\|_{\ell_2} \geq \frac{1}{N^t} \left( \|p_t\|_{\ell_2} - \left( \|p_{t+1}\|_{\ell_2} \frac{(r_t)_{t+1}}{N} + \dots + \|p_l\|_{\ell_2} \frac{(r_t)_l}{N^{l-t}} \right) \right) \geq \frac{1}{N^t} \left( \|p_t\|_{\ell_2} - \frac{k2^k}{N} \right),$$

and

$$\|q_t\|_{\ell_2}^2 \geq \frac{1}{N^{2t}} \left( \|p_t\|_{\ell_2}^2 - \frac{2k2^k}{N} \right).$$

Finally,

$$\begin{aligned}
\mathbb{E}[p^2] &= \sum_{t=0}^l \mathbb{E}[q_t^2] \\
&\geq \sum_{t=0}^l \sigma(m-1, l-t) \cdot \|q_t\|_{\ell_2}^2 \\
&\geq \sum_{t=0}^l \sigma(m-1, l-t) \cdot \frac{1}{N^{2t}} \left( \|p_t\|_{\ell_2}^2 - \frac{2k2^k}{N} \right) \\
&\geq \sum_{t=0}^l \frac{(1 - \frac{m-1}{2n})}{N^{2l-2t}} \cdot \frac{1}{N^{2t}} \cdot \left( \|p_t\|_{\ell_2}^2 - \frac{2k2^k}{N} \right) \\
&= \frac{(1 - \frac{m-1}{2n})}{N^{2l}} \cdot \sum_{t=0}^l \left( \|p_t\|_{\ell_2}^2 - \frac{2k2^k}{N} \right) \\
&\geq \frac{(1 - \frac{m-1}{2n})}{N^{2l}} \cdot \left( 1 - \frac{2K^2 2^k}{N} \right).
\end{aligned}$$

Take  $N := 4nK^2 2^k$  so that  $(1 - \frac{m-1}{2n}) \cdot (1 - \frac{2K^2 2^k}{N}) \geq 1 - \frac{m-1}{2n} - \frac{2K^2 2^k}{N} = 1 - \frac{m}{2n}$ . This completes the induction and proves the lemma.  $\blacksquare$

**Construction of  $\widehat{W}$ .** We now prove Theorem 7.4. Given  $n$  and  $q$ , let  $k = \frac{q}{2}$ , and consider random variables  $x_1, \dots, x_n$  above. Let  $\widehat{W} \in \mathbb{R}^{\mathbb{N}_k^n \times \mathbb{N}_k^n}$  be such that for any  $\alpha, \beta \in \mathbb{N}_k^n$ ,  $\widehat{W}[\alpha, \beta] = \mathbb{E}[x^{\alpha+\beta}] \cdot N^{2k}$ . By definition,  $\widehat{W}$  is degree- $q$  SoS symmetric. Since each entry of  $\widehat{W}$  corresponds to a monomial of degree exactly  $q$  and each  $x_i$  is drawn independently from the Wigner semicircle distribution, each entry of  $\widehat{W}$  is at most the  $\frac{q}{2}$ th Catalan number  $C_{\frac{q}{2}} \leq 2^q$ . For any unit vector  $p = (p_S)_{S \in \mathbb{N}_k^n} \in \mathbb{R}^{\mathbb{N}_k^n}$ , Lemma 7.7 shows  $p^T \widehat{W} p = \mathbb{E}[p^2] \cdot N^{2k} \geq \frac{1}{2}$  where  $p$  also represents a degree- $k$  homogeneous polynomial  $p(x_1, \dots, x_n) = \sum_{\alpha \in \binom{[n]}{k}} p_\alpha x^\alpha$ . Therefore, the minimum eigenvalue of  $\widehat{W}$  is at least  $\frac{1}{2}$ .

### 7.2.2. Final Construction

Let  $A \in \mathbb{R}^{[n]^{d/2} \times [n]^{d/2}}$  be a random matrix whose entry is an independent Rademacher random variable, and let  $f := \sum_{I, J \in [n]^{d/2}} A[I, J] \cdot x^{\alpha(I) + \alpha(J)}$ . Our lower bound on  $\|f\|_2$  by is proved by constructing a moment matrix  $M \in \mathbb{R}^{[n]^{d/2} \times [n]^{d/2}}$  that is

- $\text{Tr}(M) = 1$ .
- positive semidefinite.
- SoS-symmetric.
- $\langle A, M \rangle \geq \frac{n^{d/4}}{2^{O(d)} d^{d/4}}$ ,

where  $A \in \mathbb{R}^{[n]^{d/2} \times [n]^{d/2}}$  is any matrix representation of  $f$  (SoS-symmetry of  $M$  ensures  $\langle A, M \rangle$  does not depend on the choice of  $A$ ).

Let  $A$  be the SoS-symmetric matrix such that for any  $I = (i_1, \dots, i_{d/2})$  and  $J = (j_1, \dots, j_{d/2})$ ,

$$A[i, j] = \begin{cases} \frac{A[I, J]}{d!}, & \text{if } i_1, \dots, i_{d/2}, j_1, \dots, j_{d/2} \text{ are all distinct.} \\ 0 & \text{otherwise.} \end{cases}$$

We bound  $\|A\|_2$  in two steps. Let  $\hat{A}_Q \in \mathbb{R}^{\mathbb{N}_{d/2}^n \times \mathbb{N}_{d/2}^n}$  be the *quotient matrix* of  $A$  defined by

$$\hat{A}_Q[\alpha, \beta] := A[I, J] \cdot \sqrt{|\mathcal{O}(\alpha)| |\mathcal{O}(\beta)|},$$

where  $I, J \in [n]^{d/2}$  are such that  $\alpha = \alpha(I), \beta = \alpha(J)$ .

**Lemma 7.8.** *With high probability,  $\|\hat{A}_Q\|_2 \leq 2^{O(d)} \cdot \frac{n^{d/4}}{d^{d/4}}$ .*

*Proof.* Consider any  $x \in \mathbb{R}^{\mathbb{N}_{d/2}^n}$  s.t.  $\|x\| = 1$ . Since

$$\begin{aligned} x^T \hat{A}_Q x &= \sum_{\beta, \gamma \in \mathbb{N}_{d/2}^n} \hat{A}_Q[\beta, \gamma] x_\beta x_\gamma \\ &= \sum_{\beta, \gamma \in \mathbb{N}_{d/2}^n} x_\beta x_\gamma \sum_{\substack{I, J \in [n]^{d/2}: \\ \alpha(I) + \alpha(J) = \alpha + \beta}} A[I, J] \cdot \frac{\sqrt{|\mathcal{O}(\alpha)| |\mathcal{O}(\beta)|}}{|\mathcal{O}(\alpha + \beta)|} \\ &= \sum_{I, J \in [n]^{d/2}} A[I, J] \sum_{\substack{\beta, \gamma \in \mathbb{N}_{d/2}^n: \\ \alpha(I) + \alpha(J) = \alpha + \beta}} \frac{\sqrt{|\mathcal{O}(\alpha)| |\mathcal{O}(\beta)|}}{|\mathcal{O}(\alpha + \beta)|} \cdot x_\beta x_\gamma \end{aligned}$$

So  $x^T \hat{A}_Q x^T$  is a sum of independent random variables

$$\sum_{I, J \in [n]^d} A[I, J] \cdot c_{I, J}$$

where each  $A[I, J]$  is independently sampled from the Rademacher distribution and

$$c_{I, J} := \sum_{\substack{\beta, \gamma \in \mathbb{N}_{d/2}^n: \\ \beta + \gamma = \alpha(I) + \alpha(J)}} \frac{\sqrt{|\mathcal{O}(\beta)| |\mathcal{O}(\gamma)|}}{|\alpha(I) + \alpha(J)|} x_\beta x_\gamma.$$

By Cauchy-Schwarz,

$$c_{I, J}^2 \leq \left( \sum_{\beta + \gamma = \alpha(I) + \alpha(J)} \frac{|\mathcal{O}(\beta)| |\mathcal{O}(\gamma)|}{|\alpha(I) + \alpha(J)|^2} \right) \cdot \left( \sum_{\beta + \gamma} x_\beta^2 x_\gamma^2 \right) \leq \sum_{\beta + \gamma = \alpha} x_\beta^2 x_\gamma^2,$$

since there are at most  $|\mathcal{O}(\alpha)|$  choices of  $\beta$  and  $\gamma$  with  $\beta + \gamma = \alpha$ , and  $|\mathcal{O}(\beta)| \cdot |\mathcal{O}(\gamma)| \leq |\mathcal{O}(\alpha)|$ . Therefore,  $x^T \hat{A}_Q x^T$  is the sum of independent random variables that are centred and always bounded in  $[-1, +1]$ . Furthermore, the total variance is  $\sum_\alpha c_\alpha^2 \leq \sum_{\beta, \gamma} x_\beta^2 x_\gamma^2 = 1$ . The claim follows from the standard concentration bound and applying union bound over a sufficiently fine net of the unit sphere in  $|\mathbb{N}_{d/2}^n| \leq 2^{O(d)} \cdot \frac{n^{d/2}}{d^{d/2}}$  dimensions.  $\blacksquare$

**Lemma 7.9.** *For any SoS-symmetric  $A \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$ ,  $\|A\|_2 \leq \|\hat{A}_Q\|_2$ .*

*Proof.* For any  $u, v \in \mathbb{R}^{[n]^{q/2}}$  s.t.  $\|u\| = \|v\| = 1$ , we have

$$u^T A v = \sum_{I, J \in [n]^{q/2}} A[I, J] u_I v_J$$

$$\begin{aligned}
&= \sum_{I, J \in [n]^{d/2}} \frac{\hat{A}_Q[\alpha(I), \alpha(J)]}{\sqrt{|\mathcal{O}(I)| |\mathcal{O}(J)|}} \cdot u_I v_J \\
&= \sum_{\alpha, \beta \in \mathbb{N}_{q/2}^n} \frac{A[\alpha, \beta]}{\sqrt{|\mathcal{O}(\alpha)| |\mathcal{O}(\beta)|}} \langle u|_{\mathcal{O}(\alpha)}, \mathbb{1} \rangle \langle v|_{\mathcal{O}(\beta)}, \mathbb{1} \rangle \\
&= a^T \hat{A}_Q b \quad \text{where } a_\alpha := \frac{\langle u|_{\mathcal{O}(\alpha)}, \mathbb{1} \rangle}{\sqrt{|\mathcal{O}(\alpha)|}}, \quad b_\alpha := \frac{\langle v|_{\mathcal{O}(\alpha)}, \mathbb{1} \rangle}{\sqrt{|\mathcal{O}(\alpha)|}} \\
&\leq \|\hat{A}_Q\|_2 \|a\| \cdot \|b\| \\
&= \|\hat{A}_Q\|_2 \sqrt{\sum_{\alpha \in \mathbb{N}_{q/2}^n} \langle u|_{\mathcal{O}(\alpha)}, \mathbb{1} \rangle^2 / |\mathcal{O}(\alpha)|} \sqrt{\sum_{\alpha \in \mathbb{N}_{q/2}^n} \langle v|_{\mathcal{O}(\alpha)}, \mathbb{1} \rangle^2 / |\mathcal{O}(\alpha)|} \\
&\leq \|\hat{A}_Q\|_2 \sqrt{\sum_{\alpha \in \mathbb{N}_{q/2}^n} \|u|_{\mathcal{O}(\alpha)}\|^2} \sqrt{\sum_{\alpha \in \mathbb{N}_{q/2}^n} \|v|_{\mathcal{O}(\alpha)}\|^2} \quad (\text{by Cauchy-Schwarz}) \\
&\leq \|\hat{A}_Q\|_2 \|u\| \cdot \|v\| = \|\hat{A}_Q\|_2. \quad \blacksquare
\end{aligned}$$

The above two lemmas imply that  $\|A\|_2 \leq \|\hat{A}_Q\|_2 \leq 2^{O(d)} \cdot \frac{n^{d/4}}{d^{d/4}}$ . Our moment matrix  $M$  is defined by

$$M := \frac{1}{c_1} \left( \frac{1}{c_2} \cdot \frac{d^{3d/4}}{n^{3d/4}} A + \frac{W}{n^{d/2}} \right),$$

where  $W$  is the direct extension of  $\hat{W}$  constructed in Theorem 7.4 —  $W[I, J] := \hat{W}[\alpha(I), \alpha(J)]$  for all  $I, J \in [n]^{d/2}$ , and  $c_1, c_2 = 2^{\Theta(d)}$  that will be determined later.

We first consider the trace of  $M$ . The trace of  $A$  is 0 by design, and the trace of  $W$  is  $n^{d/2} \cdot 2^{O(d)}$ . Therefore, the trace of  $M$  can be made 1 by setting  $c_1$  appropriately. Since both  $A$  and  $W$  are SoS-symmetric, so is  $M$ . Since  $\mathbb{E}[W, A] = 0$  and for each  $I, J \in [n]^{d/2}$  with  $i_1, \dots, i_{d/2}, j_1, \dots, j_{d/2}$  all distinct we have  $\mathbb{E}[A[I, J] A[I, J]] = \frac{1}{d!}$ , with high probability

$$\langle A, M \rangle = \frac{1}{c_1} \cdot \left\langle A, \left( \frac{1}{c_2} \cdot \frac{d^{3d/4}}{n^{3d/4}} A + \frac{W}{n^{d/2}} \right) \right\rangle \geq 2^{O(-d)} \cdot \frac{d^{3d/4}}{n^{3d/4}} \cdot \frac{n^d}{d^d} = 2^{O(-d)} \cdot \frac{n^{d/4}}{d^{d/4}}.$$

It finally remains to show that  $M$  is positive semidefinite. Take an arbitrary vector  $v \in \mathbb{R}^{[n]^{d/2}}$ , and let

$$p = \sum_{\alpha \in \mathbb{N}_{d/2}^n} x^\alpha p_\alpha = \sum_{\alpha \in \mathbb{N}_{d/2}^n} x^\alpha \cdot \left( \sum_{I \in [n]^{d/2}: \alpha(I) = \alpha} v_I \right)$$

be the associated polynomial. If  $p = 0$ , SoS-symmetry of  $M$  ensures  $v M v^T = 0$ . Normalize  $v$  so that  $\|p\|_{\ell_2} = 1$ . First, consider another vector  $v_m \in [n]^{d/2}$  such that

$$(v_m)_I = \begin{cases} \frac{p^{\alpha(I)}}{(d/2)!}, & \text{if } i_1, \dots, i_{d/2} \text{ are all distinct.} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\|v_m\|_2^2 \leq \sum_{\alpha \in \mathbb{N}_{d/2}^n} p_\alpha^2 / (d/2)! = \frac{1}{(d/2)!},$$

so  $\|v_m\|_2 \leq \frac{2^{O(d)}}{d^{d/4}}$ . Since  $A$  is SoS-symmetric, has the minimum eigenvalue at least  $-2^{O(d)} \cdot \frac{n^{d/4}}{d^{d/4}}$ , and has nonzero entries only on the rows and columns  $(i_1, \dots, i_{d/2})$  with all different entries,

$$v^T A v = (v_m)^T A (v_m) \geq 2^{O(-d)} \cdot \frac{n^{d/4}}{d^{3d/4}}.$$

We finally compute  $v^T W v$ . Let  $v_w \in [n]^{q/2}$  be the vector where for each  $\alpha \in \mathbb{N}_{d/2}^n$ , we choose one  $I \in [n]^{q/2}$  arbitrarily and set  $(v_w)_I = p_\alpha$  (all other  $(v_w)_I$ 's are 0). By SoS-symmetry of  $W$ ,

$$v^T W v = (v_w)^T W (v_w) = p^T \hat{W} p \geq \frac{1}{2},$$

by Theorem 7.4. Therefore,

$$v M v^T = \frac{1}{c_1} \cdot v \left( \frac{1}{c_2} \cdot \frac{d^{3d/4}}{n^{3d/4}} A + \frac{W}{n^{d/2}} \right) v^T \geq \frac{1}{c_1} \cdot \left( \frac{1}{c_2} \cdot 2^{O(-d)} \cdot \frac{n^{d/4}}{d^{3d/4}} \cdot \frac{d^{3d/4}}{n^{3d/4}} + \frac{1}{2} \cdot \frac{1}{n^{d/2}} \right) \geq 0,$$

by taking  $c_2 = 2^{\Theta(d)}$ . So  $M$  is positive semidefinite, and this finishes the proof of Theorem 1.3.

## 8. Lower Bounds for Non-negative Coefficient Polynomials

Let  $G = (V, E)$  be a random graph drawn from the distribution  $G_{n,p}$ , where  $p = \frac{1}{n^3}$  and  $V = [n]$ . Let  $C_4 \subseteq \binom{V}{4}$  be the set of 4-cliques in  $G$ . The polynomial  $f$  is defined as

$$f(x_1, \dots, x_n) := \sum_{\{i_1, i_2, i_3, i_4\} \in C_4} x_{i_1} x_{i_2} x_{i_3} x_{i_4}.$$

Clearly,  $f$  is multilinear and every coefficient of  $f$  is nonnegative. The main result of this section is the following two lemmas that establish a large gap between  $\|f\|_2$  and  $\Lambda(f)$ .

**Lemma 8.1.**  $\|f\|_2 \leq n^{1/12} \log^{6.5} n$  with high probability.

**Lemma 8.2.**  $\Lambda(f) \geq \frac{n^{1/6}}{\log^2 n}$  with high probability.

### 8.1. Upper Bound on $\|f\|_2$

#### 8.1.1. Discretization

Let  $x^* \in \mathbb{S}^{n-1}$  be the vector that maximizes  $f$ . Without loss of generality, assume that every coordinate of  $x^*$  is nonnegative. Let  $y^*$  be another unit vector defined by  $y^* := (x^* + \frac{\mathbb{1}}{\sqrt{n}}) / \|x^* + \frac{\mathbb{1}}{\sqrt{n}}\|$ . Since both  $x^*$  and  $\frac{\mathbb{1}}{\sqrt{n}}$  are unit vectors, the denominator is at most 2. This implies that  $f(y^*) \geq \frac{f(x^*)}{2^4}$ , and each coordinate of  $y^*$  is at least  $\frac{1}{2\sqrt{n}}$ . For  $1 \leq j \leq \log_2 n$ , let  $Y_j \subseteq [n]$  be the set of coordinates  $i$  such that  $2^{-j} < y_i^* \leq 2^{-(j-1)}$ . The sets  $Y_1, \dots, Y_{\log_2 n}$  partition  $[n]$ . Since  $1 = \sum_{i \in [n]} y_i^2 > |Y_j| 2^{-2j}$  for each  $j$ ,  $|Y_j| \leq 2^{2j}$ .

Let  $Z_1, Z_2, Z_3$ , and  $Z_4$  be random subsets of  $[n]$  defined by the following random procedure. Let  $z_1, z_2, z_3, z_4$  be independent random variables sampled uniformly from  $\{1, \dots, \log_2 n\}$ . Also randomly partition each  $Y_i$  to  $Y_{i,1}, \dots, Y_{i,4}$  where each element of  $Y_i$  is put into exactly one of

$Y_{i,1}, \dots, Y_{i,4}$  uniformly and independently. For  $i = 1, \dots, 4$ ,  $Z_i := Y_{z_i,i}$ . Note that  $Z_1, \dots, Z_4$  are pairwise disjoint by design.

Consider a monomial  $x_{i_1}x_{i_2}x_{i_3}x_{i_4}$  where  $i_1, \dots, i_4$  form a 4-clique in  $G$ . This monomial contributes  $t := y_{i_1}^*y_{i_2}^*y_{i_3}^*y_{i_4}^*$  to  $f(y^*)$ . Let  $j_1, \dots, j_4$  be such that  $i_k \in Y_{j_k}$  for  $k = 1, \dots, 4$ . By definition of  $Y_j$ ,  $t \leq 2^{-(j_1+j_2+j_3+j_4-4)}$ . For  $k = 1, \dots, 4$ ,  $\Pr[i_k \in Z_k] = \frac{1}{4 \log n}$ . Since  $i_1, \dots, i_4$  are different, these 4 events are independent. Therefore,

$$\begin{aligned} \mathbb{E} \left[ \frac{\mathbb{I}[i_k \in Z_k \text{ for all } k = 1, \dots, 4]}{\sqrt{|Z_1||Z_2||Z_3||Z_4|}} \right] &= \prod_{k=1}^4 \mathbb{E} \left[ \frac{\mathbb{I}[i_k \in Z_k]}{\sqrt{|Z_k|}} \right] \geq \prod_{k=1}^4 \frac{1}{4 \log n \sqrt{|Y_{j_k}|}} \\ &\geq \prod_{k=1}^4 \frac{2^{-j_k}}{4 \log n} \geq \left( \frac{1}{4 \log n} \right)^4 \cdot 2^{-(j_1+j_2+j_3+j_4)} \geq \left( \frac{1}{8 \log n} \right)^4 \cdot t. \end{aligned}$$

Say that a 4-clique is *shattered* by 4 disjoint sets  $Z_1, \dots, Z_4$  if it has exactly one vertex in each of the 4 sets. By linearity of expectation,

$$\mathbb{E} \left[ \frac{\text{Number of 4-cliques shattered by } Z_1, \dots, Z_4}{\sqrt{|Z_1||Z_2||Z_3||Z_4|}} \right] \geq \left( \frac{1}{8 \log n} \right)^4 f(y^*) \geq \left( \frac{1}{16 \log n} \right)^4 f(x^*),$$

which implies that there exist four pairwise disjoint sets  $Z_1, \dots, Z_4 \subseteq V$  such that the number of shattered 4-cliques, divided by  $\sqrt{|Z_1||Z_2||Z_3||Z_4|}$ , is at least  $\frac{\|f\|_2}{(16 \log n)^4}$ . We will show that with high probability,  $G$  satisfies the property that every four disjoint sets  $Z_1, \dots, Z_4 \subseteq V$  shatter at most  $O(\sqrt{|Z_1||Z_2||Z_3||Z_4|} \cdot n^{1/12} \log^{2.5} n)$  4-cliques, proving Lemma 8.1.

### 8.1.2. Bounding Edge Density

Given a vertex  $v \in V$ , let  $\text{nbr}(v)$  be the set of vertices adjacent to  $v$ , except  $v$ . Given three different vertices  $u, v, w \in V$ , let  $\text{nbr}(u, v, w) := \text{nbr}(u) \cap \text{nbr}(v) \cap \text{nbr}(w)$ . We first prove the following lemma that three different vertices cannot share many common neighbors.

**Lemma 8.3.** *With probability at least  $1 - \frac{1}{n}$ ,  $G$  has the following property: for any three distinct vertices  $u, v, w \in V$ ,  $|\text{nbr}(u, v, w)| \leq c_1 \log n$  for some universal constant  $c_1$ .*

*Proof.* For a fixed triple  $(u, v, w)$ , a vertex  $x \in V$  is in  $\text{nbr}(u, v, w)$  with probability  $p^3 = \frac{1}{n}$ , and  $\mathbb{E}[|\text{nbr}(u, v, w)|] = (n-3) \cdot \frac{1}{n} \leq 1$ . By Chernoff bound,

$$\Pr[|\text{nbr}(u, v, w)| > c_1 \log n] \leq \exp(-4 \log n) = \frac{1}{n^4}.$$

for some constant  $c_1$ . The lemma follows by taking union bound over all triples  $(u, v, w)$ . ■

We also prove the following lemma bounding the number of edges between two disjoint sets.

**Lemma 8.4.** *With probability at least  $1 - \frac{1}{n}$ ,  $G$  has the following property: for any two disjoint sets  $Z_1, Z_2 \in V$  with  $z_1 := |Z_1|, z_2 := |Z_2|$ ,  $|E(Z_1, Z_2)| \leq c_2 \max(z_1 z_2 p, z_1 + z_2) \log n$  for some universal constant  $c_2$ .*

*Proof.* Let  $Z_1, Z_2 \subseteq V$  be two disjoint sets. Let  $z_1 := |Z_1|$  and  $z_2 := |Z_2|$ . The expected number of edges between  $Z_1$  and  $Z_2$  is  $z_1 z_2 p$ . By Chernoff bound,

$$\Pr[|E(Z_1, Z_2)| > c_2 \max(z_1 z_2 p, z_1 + z_2) \log n] \leq \exp(-4 \max(z_1 z_2 p, z_1 + z_2) \log n) \leq \frac{1}{n^{4(z_1+z_2)}}.$$

for some constant  $c_2$ . The lemma follows by taking union bound over all sets of sizes  $z_1$  and  $z_2$  (total at most  $n^{z_1+z_2}$  choices), and over all choices of  $z_1$  and  $z_2$  (at most  $n^2$  choices). ■

### 8.1.3. Bounding Triangle Density

We also study the number of triangles in  $G$  shattered by sets of size  $z_1, z_2$ , and  $z_3$  respectively and prove two results. The first result relies on counting arguments based on the guarantees provided by Lemma 8.3 and Lemma 8.4.

**Lemma 8.5.** *With probability at least  $1 - \frac{2}{n}$ ,  $G$  has the following property: for any three disjoint sets  $Z_1, Z_2, Z_3 \subseteq V$  with  $z_1 := |Z_1|, z_2 := |Z_2|, z_3 := |Z_3|$  ( $z_1 \leq z_2 \leq z_3$ ), the number of triangles shattered by  $Z_1, Z_2, Z_3$  is at most  $\max(z_{12} \sqrt{2c_1 z_3 \log n}, 2z_3)$ , where  $z_{12} \leq c_2 \max(z_1 z_2 p, z_1 + z_2) \log n$  denotes the number of edges between  $Z_1$  and  $Z_2$ .*

*Proof.* With probability at least  $1 - \frac{2}{n}$ ,  $G$  satisfies the conditions stated in Lemma 8.3 and Lemma 8.4. Fix such a  $G$ , and consider arbitrary subsets  $Z_1, Z_2, Z_3 \subseteq V$ . Let  $z_{123}$  be the number of triangles in  $G$  shattered by  $Z_1, Z_2, Z_3$ . Consider the bipartite graph  $H$  where the left side vertices correspond to edges in  $E(Z_1, Z_2)$ , the right side vertices correspond to vertices in  $Z_3$ , and there is an edge from  $(v_1, v_2) \in E(Z_1, Z_2)$  to  $v_3 \in Z_3$  when both  $(v_1, v_3), (v_2, v_3) \in E$ . Clearly,  $z_{123}$  is equal to the number of edges in  $H$ . Consider two different edges  $(v_1, v_2), (u_1, u_2) \in E(Z_1, Z_2)$ . These two edges are incident on at least 3 vertices, so the number of vertices  $w \in V$  that are adjacent to all  $\{v_1, v_2, u_1, u_2\}$  in  $G$  is at most  $c_1 \log n$ . This means that  $(v_1, v_2)$  and  $(u_1, u_2)$ , as vertices in  $H$ , share at most  $c_1 \log n$  common neighbors.

Let  $r$  be the number of the ordered triples  $((v_1, v_2), (u_1, u_2), w_3)$  where  $(v_1, v_2) \neq (u_1, u_2) \in E(Z_1, Z_2)$ ,  $w_3 \in Z_3$ , and  $w_3$  is adjacent to both  $(v_1, v_2)$  and  $(u_1, u_2)$  in  $H$ . We use the following double counting argument for  $r$  to prove the lemma.

First, since  $(v_1, v_2)$  and  $(u_1, u_2)$  share at most  $c_1 \log n$  common neighbors,  $r \leq c_1 z_{12}^2 \log n$ . On the other hand, for every vertex  $v \in V_3$ , let  $\deg_H(v)$  be its degree in  $H$ . Note that  $z_{123}$  is the number of edges in  $H$ , which is equal to  $\sum_{v \in V_3} \deg_H(v)$ . Counting  $r$  based on the right side vertices gives

$$r = \sum_{v \in V_3} \deg_H(v)(\deg_H(v) - 1) = \sum_{v \in V_3} \deg_H(v)^2 - z_{123}.$$

By Cauchy-Schwarz,

$$\sum_{v \in V_3} \deg_H(v)^2 \geq \frac{(\sum_{v \in V_3} \deg_H(v))^2}{z_3} = \frac{z_{123}^2}{z_3}.$$

Combining with  $r \leq c_1 z_{12}^2 \log n$ ,

$$c_1 z_{12}^2 \log n \geq r \geq \sum_{v \in V_3} \deg_H(v)^2 - z_{123} \geq \frac{z_{123}^2}{z_3} - z_{123}.$$

If  $z_{123} \leq \frac{z_{123}^2}{2z_3}$ , the final expression becomes at least  $\frac{z_{123}^2}{2z_3}$  and we have  $z_{123} \leq z_{12} \sqrt{2c_1 z_3 \log n}$ . Otherwise,  $z_{123} \leq 2z_3$ . ■



The second result does not rely on the previous lemmas, and directly derive the concentration bound on the number of triangles shattered by three fixed sets  $Z_1, Z_2, V \setminus (Z_1 \cup Z_2)$ . To take union bound over the choice of  $Z_1$  and  $Z_2$ , we compute sufficiently high moments to achieve a good concentration.

**Lemma 8.6.** *With probability at least  $1 - \frac{1}{n}$ ,  $G$  has the following property: for any two disjoint sets  $Z_1, Z_2 \in V$  with  $z_1 := |Z_1|, z_2 := |Z_2|$  ( $z_1 \leq z_2$ ) such that  $|z_2| \geq \sqrt{n}$ , the number of triangles shattered by  $Z_1, Z_2, V \setminus (Z_1 \cup Z_2)$  is at most  $8ez_2^2$ .*

*Proof.* Fix two disjoint sets  $Z_1, Z_2 \subseteq V$ . Let  $Z_3 := V \setminus (Z_1 \cup Z_2)$  and  $z_3 := |Z_3|$ . Let  $G_1, \dots, G_{z_1 z_2 z_3}$  be all copies of triangles in the complete tripartite graph on  $Z_1, Z_2$ , and  $Z_3$ , and  $Y_i$  be the random variable indicating whether  $G_{z_1 z_2 z_3}$  forms a triangle in  $G$  or not. Then  $Y := \sum_i Y_i$  is the random variable measuring the number of triangles shattered by  $Z_1, Z_2, Z_3$ .

**Claim 8.7.** *For  $1 \leq m \leq 3z_2 \log n$ ,  $\mathbb{E}[Y^m] \leq (8z^2)^m$ .*

*Proof.* When  $m = 1$ ,  $\mathbb{E}[Y] = z_1 z_2 z_3 p^3 \leq z_1 z_2 \leq 8z^2$ . For  $m \geq 2$ , the moments of  $Y$  can be estimated as follows, similarly to Janson et al. [JOR04] Given a graph  $H$ , let  $e(H)$  be the number of edges in  $H$ . Let  $F = F(i_1, \dots, i_{m-1}) = G_{i_1} \cup \dots \cup G_{i_{m-1}}$ . We bound the number of triangles depending on how its edges intersect those of  $F$ .

- Edge-disjoint: the number of triangles edge-disjoint from  $F$  is bounded by  $z_1 z_2 z_3$ .
- Intersecting in one edge: there are at most  $3(m-1)z_3 \leq 3(m-1)n$  triangles intersecting with  $F$  in one edge.
- Intersecting in two edges: there are at most  $3(m-1)^2$  triangles intersecting with  $F$  in two edges.
- Contained in  $F$ : there are at most  $(m-1)^{3/2}$  triangles contained in  $F$  [JOR04].

$$\begin{aligned} \mathbb{E}[Y^m] &= \sum_{i_1, \dots, i_m} p^{e(G_{i_1} \cup \dots \cup G_{i_m})} \\ &= \sum_{i_1, \dots, i_{m-1}} p^{e(F)} \sum_{i_m} p^{3-e(F \cap G_{i_m})} \\ &\leq \sum_{i_1, \dots, i_{m-1}} p^{e(F)} (z_1 z_2 z_3 p^3 + 3(m-1)np^2 + 3(m-1)^2 p + (m-1)^{3/2}). \end{aligned}$$

When  $z_2 \geq \sqrt{n}$  and  $m \leq 3z_2 \log n$ , four quantities  $z_1 z_2 z_3 p^3, 3(m-1)np^2, 3(m-1)^2 p, (m-1)^{3/2}$  are all at most  $z_2^2$ . Therefore,

$$\mathbb{E}[Y^m] \leq \sum_{i_1, \dots, i_{m-1}} p^{e(F)} \cdot 8z^2 \leq (8z^2)^{m-1} 8z^2 = (8z^2)^m,$$

completing the induction. ■

Let  $m = 3z_2 \log n$ . By Markov's inequality,

$$\Pr[Y \geq e(8z^2)] = \Pr[Y^m \geq (e(8z^2))^m] \leq e^{-m} = n^{-3z_2}.$$

Taking union bound over all choices of  $Z_1, Z_2$  (at most  $n^{z_1+z_2} \leq n^{2z_2}$  choices) and all choices of  $z_1$  and  $z_2$  (at most  $n^2$  choices) prove the lemma. ■

#### 8.1.4. Bounding 4-clique Density

With probability  $1 - \frac{3}{n}$ ,  $G$  satisfies all conditions in Lemma 8.3, Lemma 8.4, Lemma 8.5, Lemma 8.6. Fix such a  $G$ . We now bound the number of 4-cliques shattered by arbitrary four sets. Fix four disjoint sets  $Z_1, \dots, Z_4 \subseteq V$ , let  $z_i := |Z_i|$ , and assume without loss of generality that  $z_1 \leq z_2 \leq z_3 \leq z_4$ . Let  $z_{12}$  be the number of edges between  $Z_1$  and  $Z_2$ ,  $z_{123}$  be the number of triangles shattered by  $Z_1, Z_2, Z_3$ , and  $z_{1234}$  be the number of 4-cliques shattered by  $Z_1, \dots, Z_4$ .

Each 4-clique shattered by  $Z_1, Z_2, Z_3, Z_4$  can be uniquely represented by a pair  $((v_1, v_2, v_3), v_4)$ , where  $(v_1, v_2, v_3)$  is a triangle shattered by  $Z_1, Z_2, Z_3$  and  $v_4$  is a vertex in  $Z_4$  adjacent to  $v_1, v_2, v_3$ . There are at most  $z_{123}$  triangles shattered by  $Z_1, Z_2, Z_3$ , and each triangle  $(v_1, v_2, v_3)$ , by Lemma 8.3 has at most  $c_1 \log n$  vertices adjacent to all  $v_1, v_2, v_3$ . Therefore,  $z_{1234}$  is at most  $c_1 z_{123} \log n$ .

By Lemma 8.5, it is bounded by

$$c_1 \max(z_{12} \sqrt{2c_1 z_3 \log n}, 2z_3) \log n.$$

If  $2z_3 > z_{12} \sqrt{2c_1 z_3 \log n}$ , this quantity becomes

$$2c_1 z_3 \log n \leq \sqrt{z_1 z_2 z_3 z_4} \cdot 2c_1 \log n,$$

since  $z_4 \geq z_3$ . Suppose  $2z_3 \leq z_{12} \sqrt{2c_1 z_3 \log n}$ , then  $z_{1234}$  is at most

$$z_{12} \sqrt{z_3} \cdot 2^{1/2} c_1^{3/2} (\log n)^{3/2} = \max(z_1 z_2 p, z_1 + z_2) \sqrt{z_3} \cdot 2^{1/2} c_1^{3/2} c_2 (\log n)^{5/2}.$$

If  $z_1 z_2 p < z_1 + z_2$ , the quantity becomes

$$(z_1 + z_2) \sqrt{z_3} \cdot 2^{1/2} c_1^{3/2} c_2 (\log n)^{5/2} \leq \sqrt{z_1 z_2 z_3 z_4} \cdot 2^{1/2} c_1^{3/2} c_2 (\log n)^{5/2}.$$

Otherwise,

$$z_1 z_2 p \sqrt{z_3} \cdot 2^{1/2} c_1^{3/2} c_2 (\log n)^{5/2} \leq (\sqrt{z_1} p) \cdot \sqrt{z_1 z_2 z_3 z_4} \cdot 2^{1/2} c_1^{3/2} c_2 (\log n)^{5/2}.$$

Therefore, in every case, Lemma 8.5 proves that

$$z_{1234} \leq (\sqrt{z_1} p) \cdot \sqrt{z_1 z_2 z_3 z_4} \cdot c_3 (\log n)^{5/2}$$

for some universal constant  $c_3$ . Lemma 8.6 proves that if  $z_2 \geq \sqrt{n}$ ,  $z_{123} \leq 8ez_2^2$ , which implies that

$$z_{1234} \leq 8ec_1 z_2^2 \log n \leq \left(\sqrt{\frac{n}{z_1}}\right) \cdot \sqrt{z_1 z_2 z_3 z_4} \cdot c_4 \log n,$$

for some constant  $c_4$ . If we let  $z_1 = n^\alpha$ ,  $(\sqrt{z_1} p) = n^{\alpha/2-1/3}$ ,  $\sqrt{\frac{n}{z_1}} = n^{1/2-\alpha/2}$ , and

$$\max_{z_1} \min(\sqrt{z_1} p, \sqrt{\frac{n}{z_1}}) = n^{1/12},$$

when  $z_1 = n^{5/6}$ . Therefore,

$$z_{1234} \leq \sqrt{z_1 z_2 z_3 z_4} \cdot \max(c_3, c_4) \cdot n^{1/12} \log n^{5/2}.$$

Combined with the discretization step, it shows that  $f(x^*) \leq O(n^{1/12} \log^{6.5} n)$ , proving Lemma 8.1.

## 8.2. Lower Bound on $\Lambda(f)$

Recall that given a random graph  $G = ([n], E)$  drawn from the distribution  $G_{n, n^{-1/3}}$ , the polynomial  $f$  is defined as

$$f(x_1, \dots, x_n) := \sum_{\{i_1, i_2, i_3, i_4\} \in C_4} x_{i_1} x_{i_2} x_{i_3} x_{i_4},$$

where  $C_4 \subseteq \binom{[n]}{4}$  is the set of 4-cliques in  $G$ .

Let  $A \in \mathbb{R}^{[n]^2 \times [n]^2}$  be the natural matrix representation of  $24f$  —  $A[(i_1, i_2), (i_3, i_4)] = 1$  if  $i_1, \dots, i_4$  form a 4-clique in  $G$ . Otherwise,  $A[(i_1, i_2), (i_3, i_4)] = 0$ . Let  $E' \subseteq [n]^2$  be the set of ordered edges — an ordered pair  $(i_1, i_2)$  is in  $E'$  if and only if  $(i_1, i_2) \in E$  as an unordered pair. Note that  $|E'| = 2m$  where  $m$  is the number of edges in  $G$ . All nonzero entries of  $A$  is contained in the principal submatrix  $A_{E'}$ , formed by the rows and columns indexed by  $E'$ . We first give a simple proof that  $\Lambda(f) \geq \Omega(n^{1/6})$  with high probability.

**Lemma 8.8.**  $\Lambda(f) \geq \Omega(\sqrt{n^2 p^5}) = \Omega(n^{1/6})$  with high probability.

*Proof.* Consider any matrix representation  $A$  of  $24f$  and its principal submatrix  $A_{E'}$ . It is easy to observe that the Frobenious norm of  $A_{E'}$  satisfies  $\|A_{E'}\|_F^2 \geq 24|C_4|$ , minimized when  $A = A$ . Since  $\|A_{E'}\|_F \leq \sqrt{|E'|} \cdot \|A_{E'}\|_2$ , with high probability,

$$\|A\|_2 \geq \|A_{E'}\|_2 \geq \sqrt{\frac{24|C_4|}{2|E'|}} = \Omega\left(\frac{\sqrt{n^4 p^6}}{\sqrt{n^2 p}}\right) = \Omega(\sqrt{n^2 p^5}). \quad \blacksquare$$

We now prove Lemma 8.2, which says that  $\Lambda(f) \geq \frac{n^{1/6}}{\log^2 n}$  with high probability. In order to show a lower bound, we present a moment matrix  $M \in \mathbb{R}^{[n]^2 \times [n]^2}$  that is positive semidefinite, SoS-symmetric, and  $\text{Tr}(M) = 1$ , that has a large inner product  $\langle A, M \rangle \geq n^{1/6} / \log^{O(1)} n$ .

Let  $I_{E'} \in \mathbb{R}^{[n]^2 \times [n]^2}$  be such that  $I[(i_1, i_2), (i_1, i_2)] = 1$  if  $(i_1, i_2) \in E'$  and all other entries are 0. Note that  $\text{Tr}(I_{E'}) = 2m$ .

For large  $\langle A, M \rangle$ , one natural attempt is to set  $M$  to be  $A$  and fix other conditions. Let  $\lambda_{\min}$  be the minimum eigenvalue of  $A$ , which is also the minimum eigenvalue of  $A_{E'}$ . Adding  $-\lambda_{\min} \cdot I_{E'}$  to  $A$  makes it positive semidefinite, so setting

$$M \leftarrow \frac{A - \lambda_{\min} I_{E'}}{\text{Tr}(A - \lambda_{\min} I_{E'})} = \frac{A - \lambda_{\min} I_{E'}}{-2m\lambda_{\min}}$$

makes sure that  $M$  is positive semidefinite,  $\text{Tr}(M) = 1$ , and  $\langle A, M \rangle = \frac{12|C_4|}{-m\lambda_{\min}}$  (each 4-clique in  $C_4$  contributes 24). Since  $|C_4| = \Theta(n^4 p^6) = \Theta(n^2)$  and  $m = \Theta(n^2 p) = \Theta(n^{5/3})$  with high probability, if  $-\lambda_{\min} = O(np^{2.5}) = O(n^{1/6})$ ,  $\langle A, M \rangle = \Omega(n^{1/6})$  as we want.

The  $M$  does not directly work since  $M$  is not SoS-symmetric. However, the following lemma proves that this issue can be fixed by losing a factor 2 in  $\langle A, M \rangle$ .

**Lemma 8.9.** *There exists  $M$  such that is positive semidefinite,  $\text{Tr}(M) = 1$ , SoS-symmetric, and  $\langle A, M \rangle = \frac{6|C_4|}{-m\lambda_{\min}}$ .*

*Proof.* Let  $Q_{E'} \in \mathbb{R}^{[n]^2 \times [n]^2}$  be the matrix such that

$$\text{ - For } (i_1, i_2) \in E', Q_{E'}[(i_1, i_1), (i_2, i_2)] = Q_{E'}[(i_2, i_2), (i_1, i_1)] = 1.$$

- For  $i \in [n]$ ,  $Q_{E'}[(i, i), (i, i)] = \deg_G(i)$ , where  $\deg_G(i)$  denotes the degree of  $i$  in  $G$ .
- All other entries are 0.

We claim that  $I_{E'} + Q_{E'}$  is SoS-symmetric —  $(I_{E'} + Q_{E'})[(i_1, i_2), (i_3, i_4)]$  has nonzero entry if and only if  $i_1 = i_2 = i_3 = i_4$  or two different numbers  $j_1, j_2$  appear exactly twice and  $(j_1, j_2) \in E$  (in this case  $(I_{E'} + Q_{E'})[(i_1, i_2), (i_3, i_4)] = 1$ ). Since  $A$  is SoS-symmetric, so  $A - \lambda_{\min}(I_{E'} + Q_{E'})$  is also SoS-symmetric.

It is easy to see that  $Q_{E'}$  is diagonally dominant, so  $Q_{E'}$  is positive semidefinite. Since we already argued that  $A - \lambda_{\min} I_{E'}$  is positive semidefinite,  $A - \lambda_{\min}(I_{E'} + Q_{E'})$  is also positive semidefinite.  $\text{Tr}(Q_{E'}) = \sum_{i \in [n]} \deg_G(i) = 2m$ . Therefore, setting

$$M \leftarrow \frac{A - \lambda_{\min}(I_{E'} + Q_{E'})}{\text{Tr}(A - \lambda_{\min}(I_{E'} + Q_{E'}))} = \frac{A - \lambda_{\min} I_{E'}}{-4m\lambda_{\min}}$$

will yield  $M$  that is PSD, SoS-symmetric,  $\text{Tr}(M) = 1$  with

$$\langle A, M \rangle = \frac{6|C_4|}{-m\lambda_{\min}}$$

as desired. ■

Therefore, it only remains to bound  $\lambda_{\min}$ , which is the minimum eigenvalue of  $A$  and  $A_{E'}$ . Our estimate  $\lambda_{\min} = -\Theta(np^{2.5})$  is based on the following observation.  $A_{E'}$  is a  $2m \times 2m$  random matrix where each row and column is expected to have  $\Theta(n^2 p^5)$  1's (the expected number of 4-cliques an edge participates). An adjacency matrix of a random graph with average degree  $d$  has a minimum eigenvalue  $-\Theta(\sqrt{d})$ , hence the estimate  $\lambda_{\min} = -\Theta(np^{2.5})$ . Even though  $A_{E'}$  is not sampled from a typical random graph model (even  $E'$  is a random variable), the following lemma shows that this estimate is correct, proving Lemma 8.2.

**Lemma 8.10.**  $\lambda_{\min} = -O(np^{2.5} \log^2 n) = -O(n^{1/6} \log^2 n)$  with high probability.

*Proof.* Instead of  $A_{E'}$ , we directly study  $A$  to bound  $\lambda_{\min}$ . For simplicity, we consider the following matrix  $\hat{A}$ , where each row and column is indexed by an unordered pair  $(i, j) \in \binom{[n]}{2}$ , and  $\hat{A}[(i_1, i_2), (i_3, i_4)] = 1$  if and only if  $i_1, i_2, i_3, i_4$  form a 4-clique.  $A$  has only zero entries in the rows or columns indexed by  $(i, i)$  for all  $i \in [n]$ , and for two pairs  $i_1 \neq i_2$  and  $i_3 \neq i_4$ , we have

$$\hat{A}[(i_1, i_2), (i_3, i_4)] = \frac{1}{4} (A[(i_1, i_2), (i_3, i_4)] + A[(i_1, i_2), (i_4, i_3)] + A[(i_2, i_1), (i_3, i_4)] + A[(i_2, i_1), (i_4, i_3)]).$$

Therefore, the minimum eigenvalue of  $\hat{A}$  is at most  $\frac{1}{4}$  times the minimum eigenvalue of  $A$ , so it suffices to bound the minimum eigenvalue of  $\hat{A}$ .

We consider the matrix  $\hat{N}_E := \hat{A} - p^4 \hat{J}_E$ , where  $\hat{J}_E \in \mathbb{R}^{\binom{[n]}{2} \times \binom{[n]}{2}}$  is such that  $\hat{J}_E[(i_1, i_2), (i_3, i_4)] = 1$  if  $(i_1, i_2), (i_3, i_4) \in E$  and 0 otherwise. Since  $\hat{J}_E$  is a rank-1 matrix with a positive eigenvalue, the minimum eigenvalues of  $\hat{A}$  and  $\hat{N}_E$  are the same. In summary,  $\hat{N}_E$  is the following matrix.

- $\hat{N}_E[(i_1, i_2), (i_3, i_4)] = 1 - p^4$  if  $i_1, i_2, i_3, i_4$  form a 4-clique in  $G$ .
- $\hat{N}_E[(i_1, i_2), (i_3, i_4)] = -p^4$  if both  $(i_1, i_2), (i_3, i_4) \in E$  but do not form a 4-clique in  $G$ .
- All other entries are 0.

We perform by the trace method to bound  $\|\widehat{\mathbf{N}}_E\|_2$ , based on the observation that for every even  $r \in \mathbb{N}$ ,  $\|\widehat{\mathbf{N}}_E\|_2 \leq (\text{Tr}((\widehat{\mathbf{N}}_E)^r))^{1/r}$ . Fix an even  $r \in \mathbb{N}$ . The expected value of the trace can be represented as

$$\mathbb{E}[\text{Tr}((\widehat{\mathbf{N}}_E)^r)] = \mathbb{E}\left[\sum_{I^1, \dots, I^r \in \binom{[n]}{2}} \prod_{k=1}^r \widehat{\mathbf{N}}_E[I^k, I^{k+1}]\right] = \sum_{I^1, \dots, I^r \in \binom{[n]}{2}} \mathbb{E}\left[\prod_{k=1}^r \widehat{\mathbf{N}}_E[I^k, I^{k+1}]\right]$$

where each  $I^j = (i_1^j, i_2^j) \in \binom{[n]}{2}$  is an edge of the complete graph on  $n$  vertices (call it a *potential edge*) and  $I^{r+1} := I^1$ .

Fix  $r$  potential edges  $I^1, \dots, I^r$ , let  $t := \prod_{k=1}^r \widehat{\mathbf{N}}_E[I^k, I^{k+1}]$ , and consider  $\mathbb{E}[t]$ . Let  $E_0 := \{I^1, \dots, I^r\}$  be the set of distinct edges represented by  $I^1, \dots, I^r$ . First, the expected value is 0 if one of  $I^j$  does not become an edge. Therefore,  $\mathbb{E}[t] = p^{|E_0|} \cdot \mathbb{E}[t | E_0 \subseteq E]$ .

Let  $D \subseteq [r]$  be the set of  $j \in [r]$  such that  $|\{i_1^j, i_2^j, i_1^{j+1}, i_2^{j+1}\}| = 4$  (i.e., two edges  $I^j, I^{j+1}$  together cover four vertices). For  $j \in [r] \setminus D$ ,  $\{i_1^j, i_2^j, i_1^{j+1}, i_2^{j+1}\}$  cannot form a 4-clique, so given that  $I^j, I^{j+1} \in E$ ,  $\widehat{\mathbf{N}}_E[I^j, I^{j+1}] = -p^4$ . For  $j \in D$ ,  $E_j := \{(i_1^j, i_1^{j+1}), (i_1^j, i_2^{j+1}), (i_2^j, i_1^{j+1}), (i_2^j, i_2^{j+1})\} \setminus E_0$  be the set of edges in the 4-clique created by  $\{i_1^j, i_2^j, i_1^{j+1}, i_2^{j+1}\}$  except ones in  $E_0$ . Then

$$\mathbb{E}[t] = p^{|E_0|} \cdot \mathbb{E}[t | E_0 \subseteq E] = p^{|E_0|} \cdot (-p^4)^{r-|D|} \cdot \mathbb{E}\left[\prod_{k \in D} \widehat{\mathbf{N}}_E[I^k, I^{k+1}] \mid E_0 \subseteq E\right].$$

Suppose there exists  $j \in D$  such that  $E_j$  does not intersect with  $E_0 \cup (\cup_{j' \in D \setminus \{j\}} E_{j'})$ . In particular,  $|E_j| = 4$ . Then, given that  $E_0 \subseteq E$ ,  $\widehat{\mathbf{N}}_E[I^j, I^{j+1}]$  becomes independent of all  $\{\widehat{\mathbf{N}}_E[I^k, I^{k+1}]\}_{k \in D \setminus \{j\}}$ , and  $\mathbb{E}[\widehat{\mathbf{N}}_E[I^j, I^{j+1}] | E_0 \subseteq E] = p^4(1 - p^4) + (1 - p^4)(-p^4) = 0$ . Therefore,  $\mathbb{E}[t] = 0$  unless for all  $j \in D$ , either  $|E_j| \leq 3$  or there exists  $j' \in D \setminus \{j\}$  with  $E_j \cap E_{j'} \neq \emptyset$ .

Let  $E' := \sum_{j \in D} E_j$ . Note that  $E_0$  and  $E'$  completely determines  $t$ .  $\mathbb{E}[t]$  can be written as

$$\begin{aligned} & p^{|E_0|} \cdot (-p^4)^{r-|D|} \cdot \mathbb{E}\left[\prod_{k \in D} \widehat{\mathbf{N}}_E[I^k, I^{k+1}] \mid E_0 \subseteq E\right] \\ &= p^{|E_0|} \cdot (-p^4)^{r-|D|} \cdot \sum_{F' \subseteq E'} \left( p^{|F'|} (1-p)^{|E'| - |F'|} \cdot \mathbb{E}\left[\prod_{k \in D} \widehat{\mathbf{N}}_E[I^k, I^{k+1}] \mid E_0 \subseteq E, E' \cap E = F'\right] \right) \\ &= p^{|E_0|} \cdot (-p^4)^{r-|D|} \cdot \sum_{F' \subseteq E'} \left( p^{|F'|} (1-p)^{|E'| - |F'|} \cdot (1-p^4)^{|D| - a(F')} (-p^4)^{a(F')} \right), \end{aligned}$$

where  $a(F')$  be the number of  $j \in D$  with  $E_j \not\subseteq F'$ . Since  $E' \subseteq F' \cup (\cup_{j: E_j \not\subseteq F'} E_j)$  and  $4a(F') + |F'| \geq |E'|$ , the expression is upper bounded by

$$p^{|E_0|} \cdot (p^4)^{r-|D|} \cdot 2^{|E'|} p^{|E'|} \leq 2^{4r} p^{4(r-D) + |E_0| + |E'|}.$$

Fix a graph  $H$  with  $r$  labelled edges  $I^1, \dots, I^r$  (possibly repeated) and  $q := q(H)$  vertices, without any isolated vertex (so  $q \leq 2r$ ). There are at most  $\binom{q}{2}^r \leq (2r)^{2r}$  such graphs. Then  $I^1, \dots, I^r$ , as edges in  $\binom{[n]}{2}$ , are determined by a map  $V_H \rightarrow [n]$ . There are at most  $n^q$  such mappings. Let  $E_0 := E_0(H)$ ,  $D := D(H)$ ,  $E_j := E_j(H)$ ,  $E' := E'(H)$  be defined as before. Note that  $E_0$  is set the edges of  $H$ . As observed before, the contribution from  $H$  is 0 if there exists  $j \in D$  such that  $|E_j| = 4$  and  $E_j$  is disjoint from  $\{E_{j'}\}_{j' \in D \setminus \{j\}}$ . Let  $\mathcal{H}$  be the set of  $H$  that has nonzero contribution.

Then  $\mathbb{E}[\text{Tr}((\hat{N}_E)^r)]$  is upper bounded by

$$\begin{aligned}
\mathbb{E}[\text{Tr}((\hat{N}_E)^r)] &= \sum_{I^1, \dots, I^r \in \binom{[n]}{2}} \mathbb{E} \left[ \prod_{k=1}^r \hat{N}_E[I^k, I^{k+1}] \right] \\
&\leq \sum_{H \in \mathcal{H}} n^{q(H)} 2^{4r} p^{4(r-D(H)) + |E_0(H)| + |E'(H)|} \\
&\leq (2r)^{2r} \max_{H \in \mathcal{H}} (n^{q(H)} 2^{4r} p^{4(r-D(H)) + |E_0(H)| + |E'(H)|}) \\
&\leq (8r)^{2r} \max_{H \in \mathcal{H}} (n^{q(H)} p^{4(r-D(H)) + |E_0(H)| + |E'(H)|}) \\
&= (8r)^{2r} \max_{H \in \mathcal{H}} (n^{q(H) - (4(r-D(H)) + |E_0(H)| + |E'(H)|)/3}).
\end{aligned}$$

The following claim bounds  $q(H) - (4(r - D(H)) + |E_0(H)| + |E'(H)|)/3$ .

**Claim 8.11.** *Any graph  $H \in \mathcal{H}$  satisfies*

$$q(H) \leq 2 + \frac{4(r - D(H)) + |E_0(H)| + |E'(H)|}{3} + \frac{r}{6}.$$

*Proof.* Fix a graph  $H \in \mathcal{H}$ . Let  $j = 1, \dots, r$ , let  $V_j := \{i_1^j, i_2^j\}_{1 \leq j \leq r}$  (i.e., the set of vertices covered by  $I^1, \dots, I^j$ ). For each  $j = 2, \dots, r$ , let  $v_j := |V_j| - |V_{j-1}|$  and classify the index  $j$  to one of the following types.

- Type  $-1$ :  $I^j$  shares at least one vertex with  $I^{j-1}$ .
- Type  $k$  ( $0 \leq k \leq 2$ ):  $I^j$  and  $I^{j-1}$  are disjoint, and  $v_j = k$  (i.e., adding  $I^j$  introduces  $k$  new vertices).

Let  $T_k$  ( $-1 \leq k \leq 2$ ) be the set of indices of Type  $k$ , and let  $t_k := |T_k|$ . The number of vertices  $q$  is bounded by

$$q \leq 2 + 1 \cdot t_{-1} + 0 \cdot t_0 + 1 \cdot t_1 + 2 \cdot t_2 = 2 + t_{-1} + t_1 + 2t_2.$$

Let  $H_j$  be the graph with  $V_j$  as vertices and  $\{i_k\}_{1 \leq k \leq j} \cup (\cup_{1 \leq k \leq j-1} E_j)$  as edges. For  $j = 2, \dots, r$ , let  $e_j$  be the number of edges in  $H_j$  minus the number of edges in  $H_{j-1}$ . The total number of edges of  $H_j$  is at least  $e_2 + \dots + e_r$ . For an index  $j \in T_2$ , adding two vertices  $i_1^j, i_2^j$  introduces at least 5 edges in  $H_j$  compared to  $H_{j-1}$  (i.e., six edges in the 4-clique on  $\{i_1^{j-1}, i_2^{j-1}, i_1^j, i_2^j\}$  except  $I^{j-1}$ ), so  $e_j \geq 5$ . The similar argument gives  $e_j \geq 3$  for each  $j \in T_1$ .

The lemma is proved via the following charging argument. For each index  $j = 2, \dots, r$ , we get value  $1/3$  for each edge in  $H_j \setminus H_{j-1}$  and get value  $1/6$  for the new index. If  $j \notin D$ , we get a value of  $4/3$ . We give this value to vertices in  $V_{j-1} \setminus V_j$ . If we do not give more value than we get and each vertex in  $V(H) \setminus V_1$  gets more than 1, this means

$$q - 2 \leq \frac{|E_0| + |E_H|}{3} + \frac{4}{3}(r - D) + \frac{r}{6},$$

proving the lemma. For example, if  $j$  is an index of Type 1, it gets a value at least  $7/6$  ( $3 \cdot 1/3 = 1$  from new edges and  $1/6$  for the index) and need to give value 1, such a charging can be done. However, if  $j$  is an index of Type 2, we can get as small as  $5/3 + 1/6 = 11/6$ , while we need to give value 2. We fix this problem by performing the distribution for such Type 2 indices with indices of other Types.

Consider an index  $j \in T_2$ . The fact that  $j \in T_2$  guarantees that earlier edges  $I^1, \dots, I^{j-1}$  are all vertex disjoint from  $I^j$ . If later edges  $I^{j+1}, \dots, I^r$  are all vertex disjoint from  $I^j$ ,  $|E_j| = 4$  and  $E_j$  is disjoint from  $\{E_{j'}\}_{j' \in D \setminus \{j\}}$ , and this means that  $H \notin \mathcal{H}$ . Since we assume  $H \in \mathcal{H}$ , there must be another index  $j' > j$  such that  $I^{j'}$  and  $I^j$  share a vertex. Take the smallest  $j' > j$ , and say that  $j'$  *compensates*  $j$ . Note that Type 2 indices cannot compensate another Type 2 index, Type 1 indices compensate at most one Type 2 index, and Type  $-1$  and Type 0 indices can compensate at most two Type 2 indices. Thus, there are six different types of compensators. For each type, we check whether the collective distribution scheme works so that we give each vertex value 1.

1. One Type 1 index  $j'$  compensates one Type 2 index  $j$ :

$$v_{j'} + v_j = 3 \text{ and } e_{j'} + e_j \geq 8 \text{ (5 from } e_j \text{ and 3 from } e_{j'}\text{). Distribution works since } 3 \leq 8/3 + 2/6.$$

2. One Type 0 index  $j'$  compensates one Type 2 index  $j$ :

$$v_{j'} + v_j = 2 \text{ and } e_{j'} + e_j \geq 5 \text{ (5 from } e_j\text{). Distribution works since } 2 \leq 5/3 + 2/6$$

3. One Type 0 index  $j'$  compensates two Type 2 indices  $j_1$  and  $j_2$ : there are two cases.

- (a)  $e_{j'} + e_{j_1} + e_{j_2} \geq 11$ :  $v_{j'} + v_{j_1} + v_{j_2} = 4$ . Distribution works since  $4 \leq 11/3 + 3/6$ .

- (b)  $e_{j'} + e_{j_1} + e_{j_2} = 10$ : since  $e_{j_1}, e_{j_2} \geq 5$ , it means that  $e_{j'} = 0$ . By the definition of  $j'$ , in  $H_{j'-1}$ , for  $k = 1, 2$ , vertices in  $I^{j_k}$  only have edges between  $I^{j_k-1}$  and  $I^{j_k+1}$ . Let  $I^{j'} = (i_1^{j'}, i_2^{j'})$  such that  $i_1^{j'} \in I^{j_1}$  and  $i_2^{j'} \in I^{j_2}$ . Since  $e_{j'} = 0$  means that  $I^{j'}$  was in  $H_{j'-1}$ , which implies that  $j_2 = j_1 + 1$  and  $j' > j_2 + 1$ . Consider  $I^{j'-1}$ , which are vertex disjoint from both  $I^{j_1}$  and  $I^{j_2}$ . If  $I^{j'-1} \neq I^{j_1-1}$ , at least one edge between  $I^{j'-1}$  and  $I_1^{j'}$  was not in  $V_{j'-1}$ , contradicting the assumption  $e_{j'} = 0$ . Therefore,  $I^{j'-1} = I^{j_1-1}$ . For the same reason,  $I^{j'-1} = I^{j_2+1}$ , which implies that  $I^{j'-1}$  is another Type 0 edge that does not compensate anyone. In this case we consider that  $I^{j'-1}$  and  $I^{j'}$  jointly compensate  $j_1$  and  $j_2$ .  $v_{j'} + v_{j''} + v_{j_1} + v_{j_2} = 4$  and  $e_{j''} + e_{j'} + e_{j_1} + e_{j_2} \geq 10$ . Distribution works since  $4 \leq 10/3 + 4/6$ .

4. One Type  $-1$  index  $j'$  compensates one Type 2 index  $j$ :

$$v_{j'} + v_j \leq 3 \text{ and } e_{j'} + e_j \geq 5 \text{ (5 from } e_j\text{). Distribution works since } 2 \leq 5/3 + 4/3 + 2/6.$$

5. One Type  $-1$  index  $j'$  compensates two Type 2 indices  $j_1$  and  $j_2$ :

$$v_{j'} + v_{j_1} + v_{j_2} \leq 5 \text{ and } e_{j'} + e_{j_1} + e_{j_2} \geq 10 \text{ (5 from } e_{j_1} \text{ and 5 from } e_{j_2}\text{). Distribution works since } 5 \leq 10/3 + 4/3 + 3/6.$$

Each index of Type 2 participates exactly one of the above collective distribution. Indices of other Type participate at most one of the above collective distribution. For indices that did not participate above, they perform an individual distribution. We finally check whether it works.

Type  $-1$ : get value at least  $3/4 + 1/6$ , and need to give value at most 1.

Type 0: get value at least  $1/6$ , and do not need to give any value.

Type 1: get value at least  $3/3$ , and need to give value at most 1.



Therefore, our distribution scheme collectively gets value at most  $\frac{|E_0|+|E'|}{3} + \frac{4}{3}(r-D) + \frac{r}{6}$  and gives value at least  $q-2$ . This proves the claim. ■

Finally,  $\mathbb{E}[\text{Tr}((\hat{N}_E)^r)]$  is bounded by

$$\begin{aligned}\mathbb{E}[\text{Tr}((\hat{N}_E)^r)] &\leq (8r)^{2r} \max_{H \in \mathcal{H}} (n^{q(H)-(4(r-D(H))+|E_0(H)|+|E'(H)|)/3}) \\ &\leq (8r)^{2r} n^{r/6+2},\end{aligned}$$

and  $\text{Tr}((\hat{N}_E)^r) \leq (8r)^{2r} n^{r/6+3}$  with probability  $1 - \frac{1}{n}$  (we indeed proved  $\mathbb{E}[\text{Tr}((\hat{N}_E)^r)]$ , so we can apply Markov's inequality). This implies that  $\|\hat{N}_E\|_2 \leq (\text{Tr}((\hat{N}_E)^r))^{1/r} \leq (8r)^2 n^{1/6+3/r}$ . Setting  $r = \log n$  gives  $\|\hat{N}_E\|_2 \leq O(n^{1/6} \log^2 n)$ , proving the lemma. ■

## References

- [BBH<sup>+</sup>12] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326. ACM, 2012. [1](#), [2](#)
- [BCK15] Boaz Barak, Siu On Chan, and Pravesh Kothari. Sum of squares lower bounds from pairwise independence. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing, STOC '15*, pages 97–106, 2015. [12](#)
- [BGK<sup>+</sup>01] Andreas Brieden, Peter Gritzmann, Ravindran Kannan, Victor Klee, László Lovász, and Miklós Simonovits. Deterministic and randomized polynomial-time approximation of radii. *Mathematika*, 48(1-2):63–105, 2001. [4](#), [5](#), [54](#), [55](#)
- [BGL16] Vijay Bhattiprolu, Venkatesan Guruswami, and Euiwoong Lee. Certifying random polynomials over the unit sphere via sum of squares hierarchy. *arXiv preprint arXiv:1605.00903*, 2016. [10](#)
- [BH13] Fernando GSL Brandao and Aram W Harrow. Quantum de finetti theorems under local measurements with applications. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 861–870. ACM, 2013. [1](#), [2](#)
- [BHK<sup>+</sup>16] Boaz Barak, Samuel B Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. In *Proceedings of the 57th annual IEEE symposium on Foundations of Computer Science, FOCS '16*, pages 428–437, 2016. [12](#)
- [BKS14] Boaz Barak, Jonathan A Kelner, and David Steurer. Rounding sum-of-squares relaxations. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 31–40. ACM, 2014. [1](#), [2](#), [5](#), [14](#)
- [BKS15] Boaz Barak, Jonathan A Kelner, and David Steurer. Dictionary learning and tensor decomposition via the sum-of-squares method. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 143–151. ACM, 2015. [1](#)

- [BS14] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *arXiv preprint arXiv:1404.5236*, 2014. 14
- [BV09] S Charles Brubaker and Santosh S Vempala. Random tensors and planted cliques. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 406–419. Springer, 2009. 1
- [Cha12] Sourav Chatterjee. The missing log in large deviations for triangle counts. *Random Structures & Algorithms*, 40(4):437–451, 2012. 13
- [Cha13] Siu On Chan. Approximation resistance from pairwise independent subgroups. In *Proceedings of the 45th annual ACM Symposium on Theory of Computing, STOC '13*, pages 447–456, 2013. 12
- [DK08] Etienne De Klerk. The complexity of optimizing over a simplex, hypercube or sphere: a short survey. *Central European Journal of Operations Research*, 16(2):111–125, 2008. 1, 4
- [DK12a] Bobby DeMarco and Jeff Kahn. Upper tails for triangles. *Random Structures & Algorithms*, 40(4):452–459, 2012. 13
- [DK12b] Robert DeMarco and Jeff Kahn. Tight upper tail bounds for cliques. *Random Structures & Algorithms*, 41(4):469–487, 2012. 13
- [DKLP06] Etienne De Klerk, Monique Laurent, and Pablo A Parrilo. A ptas for the minimization of polynomials of fixed degree over the simplex. *Theoretical Computer Science*, 361(2):210–225, 2006. 1
- [dKLS14] Etienne de Klerk, Monique Laurent, and Zhao Sun. Convergence analysis for lasserre’s measure-based hierarchy of upper bounds for polynomial optimization. *arXiv preprint arXiv:1411.6867*, 2014. 2
- [DW12] Andrew C Doherty and Stephanie Wehner. Convergence of sdp hierarchies for polynomial optimization on the hypersphere. *arXiv preprint arXiv:1210.5048*, 2012. 2, 4
- [FK08] Alan Frieze and Ravi Kannan. A new approach to the planted clique problem. In *LIPICs-Leibniz International Proceedings in Informatics*, volume 2. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2008. 1
- [FV<sup>+</sup>62] David G Feingold, Richard S Varga, et al. Block diagonally dominant matrices and generalizations of the gerschgorin circle theorem. *Pacific J. Math*, 12(4):1241–1250, 1962. 27
- [GM15] Rong Ge and Tengyu Ma. Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, page 829, 2015. 1
- [Hås96] Johan Håstad. Clique is hard to approximate within  $n^{1-\epsilon}$ . In *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pages 627–636. IEEE, 1996. 1

- [HKP<sup>+</sup>16a] Samuel B. Hopkins, Pravesh Kothari, Aaron Henry Potechin, Prasad Raghavendra, and Tselil Schramm. On the integrality gap of degree-4 sum of squares for planted clique. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '16, pages 1079–1095, Philadelphia, PA, USA, 2016. Society for Industrial and Applied Mathematics. 12
- [HKP16b] Samuel B Hopkins, Pravesh K Kothari, and Aaron Potechin. *Personal communication*, 2016. 5
- [HLZ10] Simai He, Zhenning Li, and Shuzhong Zhang. Approximation algorithms for homogeneous polynomial optimization with quadratic constraints. *Mathematical Programming*, 125(2):353–383, 2010. 1, 2, 3, 4, 7, 20
- [HSS15] Samuel B Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In *Proceedings of The 28th Conference on Learning Theory*, pages 956–1006, 2015. 1, 5
- [HV04] Johan Håstad and Srinivasan Venkatesh. On the advantage over a random assignment. *Random Structures & Algorithms*, 25(2):117–149, 2004. 1
- [JOR04] Svante Janson, Krzysztof Oleszkiewicz, and Andrzej Ruciński. Upper tails for subgraph counts in random graphs. *Israel Journal of Mathematics*, 142(1):61–92, 2004. 13, 44
- [KN08] Subhash Khot and Assaf Naor. Linear equations modulo 2 and the  $\ell_1$  diameter of convex bodies. *SIAM Journal on Computing*, 38(4):1448–1463, 2008. 5, 7, 54, 55
- [KV04] Jeong Han Kim and Van H Vu. Divide and conquer martingales and the number of triangles in a random graph. *Random Structures & Algorithms*, 24(2):166–174, 2004. 13
- [Las09] Jean Bernard Lasserre. *Moments, positive polynomials and their applications*, volume 1. World Scientific, 2009. 1, 4
- [Lau09] Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, pages 157–270. Springer, 2009. 1, 2, 4, 14
- [LZ16] Eyal Lubetzky and Yufei Zhao. On the variational problem for upper tails in sparse random graphs. *Random Structures & Algorithms*, 2016. 13
- [MPW15] Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares lower bounds for planted clique. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, pages 87–96. ACM, 2015. 12
- [MR14] Andrea Montanari and Emile Richard. A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems*, pages 2897–2905, 2014. 1, 5
- [Nes03] Yurii Nesterov. Random walk in a simplex and quadratic optimization over convex polytopes. Technical report, UCL, 2003. 1
- [OZ13] Ryan O’Donnell and Yuan Zhou. Approximability and proof complexity. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1537–1556. Society for Industrial and Applied Mathematics, 2013. 14

- [RRS16] Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random csp's below the spectral threshold. *arXiv preprint arXiv:1605.00058*, 2016. [5](#)
- [Sch08] G. Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *Proceedings of the 49th annual IEEE symposium on Foundations of Computer Science, FOCS '08*, pages 593–602, Oct 2008. Newer version available at the author's homepage. [12](#)
- [Sim98] Barry Simon. The classical moment problem as a self-adjoint finite difference operator. *Advances in Mathematics*, 137(1):82–203, 1998. [36](#)
- [So11] Anthony Man-Cho So. Deterministic approximation algorithms for sphere constrained homogeneous polynomial optimization problems. *Mathematical programming*, 129(2):357–382, 2011. [1](#), [2](#), [3](#), [4](#), [5](#), [7](#), [54](#)
- [TS14] R. Tomioka and T. Suzuki. Spectral norm of random tensors. *ArXiv e-prints*, July 2014. [4](#), [10](#)
- [Tul09] Madhur Tulsiani. CSP gaps and reductions in the Lasserre hierarchy. In *Proceedings of the 41st annual ACM Symposium on Theory of Computing, STOC '09*, pages 303–312, 2009. [12](#)
- [Vu01] Van H Vu. A large deviation result on the number of small subgraphs of a random graph. *Combinatorics, Probability & Computing*, 10(1):79–94, 2001. [13](#)

## A. Oracle Lower Bound

Khot and Naor [KN08] observed that the problem of maximizing a polynomial over unit sphere can be reduced to computing diameter of centrally symmetric convex body. This observation was also used by So [So11] later. We recall the reduction here: For a convex set  $K$ , let  $K^\circ$  denote the polar of  $K$ , i.e.,  $K^\circ = \{y : \forall x \in K \langle x, y \rangle \leq 1\}$ . For a degree-3 polynomial  $P(x, y, z)$  on  $3n$  variables, let  $\|x\|_P = \|P(x, \cdot, \cdot)\|_{s_P}$  where  $P(x, \cdot, \cdot)$  is a degree-2 restriction of  $P$  with  $x$  variables set. Let  $\mathbb{B}_P = \{x : \|x\|_P \leq 1\}$ . From the definition of polar and  $\|\cdot\|_{s_P}$ , we have:

$$\begin{aligned}
& \max_{\|x\|_2, \|y\|_2, \|z\|_2 \leq 1} P(x, y, z) \\
&= \max_{x \in \mathbb{B}_2} \|x\|_P \\
&= \max_{x \in \mathbb{B}_P^\circ} \|x\|_2
\end{aligned}$$

For general convex bodies, a lower bound for number of queries with “weak separation oracle” for approximating the diameter of the convex body was proved by Brieden et al. [BGK<sup>+</sup>01] and later improved by Khot and Naor [KN08]. We recall the definition:

**Definition A.1.** For a given a convex body  $P$ , a weak separation oracle  $A$  is an algorithm which on input  $(x, \varepsilon)$  behaves as following:

- If  $x \in A + \varepsilon \mathbb{B}_2$ ,  $A$  accepts it.
- Else  $A$  outputs a vector  $c \in \mathbb{Q}^n$  with  $\|c\|_\infty = 1$  such that for all  $y$  such that  $y + \varepsilon \mathbb{B}_2 \subset P$  we have  $c^T x + \varepsilon \geq c^T y$ .

Let  $K_{s,v}$  be the convex set  $K_{s,v}^{(n)} = \text{conv}(\mathbb{B}^n \cup \{sv, -sv\})$ , for unit vector  $v$ . Brieden et al. [BGK<sup>+</sup>01] proved the following theorem:

**Theorem A.2.** *Let  $A$  be a randomized algorithm, for every convex set  $P$ , with access to a weak separation oracle for  $P$ . Let  $\mathcal{K}(n, s) = \{K_{s,u}^{(n)}\}_{u \in \mathbb{B}_2^n} \cup \{\mathbb{B}_2^n\}$ . If for every  $K \in \mathcal{K}(n, s)$  and  $s = \frac{\sqrt{n}}{\lambda}$ , we have:*

$$\Pr \left[ A(K) \leq \text{diam}(K) \leq \frac{\sqrt{n}}{\lambda} A(K) \right] \geq \frac{3}{4}$$

where  $\text{diam}(K)$  is the diameter of  $K$ , then  $A$  must use at least  $O(\lambda 2^{\lambda^2/2})$  oracle queries for  $\lambda \in [\sqrt{2}, \sqrt{n/2}]$ .

Using  $\lambda = \log n$ , we get that to get  $s = \frac{\sqrt{n}}{\log n}$  approximation to diameter,  $A$  must use super-polynomial number of queries to the weak separation oracle. We note that this was later improved to give analogous lower bound on the number of queries for an approximation factor  $\sqrt{\frac{n}{\log n}}$  by Khot and Naor [KN08].

Below, we show that the family of hard convex bodies considered by Brieden et al. [BGK<sup>+</sup>01] can be realized as  $\{\mathbb{B}_P^\circ\}_{P \in \mathcal{P}}$  by a family of polynomials  $\mathcal{P}$  – which, in turn, establishes a lower bound of  $\Omega\left(\frac{\sqrt{n}}{\log n}\right)$  on the approximation for polynomial optimization, achievable using this approach, for the case of  $d = 3$ . Consider the degree-3 multi-linear polynomial

$$P(x, y, z) = \sum_{i=1}^n x_i y_i z_1 + s \cdot x_1 y_n z_n$$

where  $x, y, z$  are  $n$ -tuple of variables. For a linear unitary map  $U : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , let  $P_U$  be the polynomial defined as:  $P_U(x, y, z) = P(Ux, Uy, Uz)$ .

Now we prove:  $\|x\|_P = \|P(x, \cdot, \cdot)\|_{sp} = \max\{\|\bar{x}\|_2, s|x_1|\}$ , where  $\bar{x} = (x_1, \dots, x_{n-1})$ , i.e., projection of  $x$  to first  $n-1$  co-ordinates. A matrix representation of  $P(x, \cdot, \cdot)$  (with rows indexed by  $y$  and columns indexed by  $z$  variables):

$$A = \begin{pmatrix} x_1 & 0 & \dots & 0 & 0 \\ x_2 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n-1} & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & s \cdot |x_1| \end{pmatrix} \text{ and so, } A^T A = \begin{pmatrix} \|\bar{x}\|_2^2 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & s^2 \cdot |x_1|^2 \end{pmatrix}.$$

When  $\|x\|_P = \|P(x, \cdot, \cdot)\|_{sp} = \|A\|_{sp} = \max\{\|\bar{x}\|_2, s|x_1|\}$ .

Let  $B_0 = \{x : \|\bar{x}\|_2 \leq 1\}$  and  $B_1 = \{x : s \cdot |x_1| \leq 1\}$ . Hence  $B_0^\circ = \{y \in \mathbb{R}^n : \|y\|_2 \leq 1, y_{n-1} = 0\}$  – since  $n$ -th co-coordinate of  $B_0$  is unbounded. Similarly,  $B_1^\circ = \{y \in \mathbb{R}^n : |y_1| \leq 1, y_2 = \dots = y_n = 0\}$ .

We note that  $\mathbb{B}_P = B_0 \cap B_1$ , and moreover,  $\mathbb{B}_{P_U} = U^{-1} \mathbb{B}_P$ . So  $\mathbb{B}_{P_U}^\circ = U((B_0 \cap B_1)^\circ) = U(\text{conv}(B_0^\circ \cup B_1^\circ))$ , where the second equality follows by De Morgan's law of polars.

For a  $v \in \mathbb{B}_2^n$ , let  $V$  be a unitary map with  $V(e_1) = v$ , where  $e_1$  is the first standard basis vector. Then, we observe:  $K_{s,v}^{(n-1)} = \mathbb{B}_{P_V}^\circ$ .

Hence for polynomial  $Q \in \mathcal{P} = \{P_U\}_U \cup \{\sum_{i=1}^n x_i y_i z_1\}$ , no randomized polynomial can approximate  $\text{diam } \mathbb{B}_Q$  within factor  $\frac{\sqrt{n}}{q}$  without using more than  $2^{\Omega(q)}$  number of queries. Since the algorithm of Khot and Naor [KN08] reduces the problem of optimizing polynomial  $Q$  to computing  $\text{diam}(\mathbb{B}_Q)$ ,  $\mathcal{P}$  shows that their analysis is almost tight.